P1 1175588

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

May 27, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/467,491*
FILING DATE: *April 30, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/09215*

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

N. WOODSON
Certifying Officer

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

05-05-03 60467491 .043003 #1

JC769 U.S. PTO 04/30/03

Atty. Docket No.: PF-1506 P	Type a plus sign (+) inside this box ➡	+

Certificate of Mailing

"Certificate of Express Mail" mailing label number EV 330 964 029 US. I hereby certify that this document and referenced attachments are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR § 1.10, addressed to: Mail Stop: Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on April 30, 2003.

By: Nancy Ramos Printed: Nancy Ramos

JC715 U.S. PTO 60/467491

PROVISIONAL APPLICATION FOR PATENT COVER SHEET
(Large Entity)

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2).

TITLE OF THE INVENTION (500 characters max)	
KINASES AND PHOSPHATASES	
INVENTOR(S)/APPLICANT(S)	
FIRST NAME/MIDDLE INITIAL/LAST NAME	**RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)**
Vicki S. Elliott	San Jose, California
Reena Khare	Saratoga, California
Thomas W. Richardson	Redwood City, California
Joseph P. Marquis	San Jose, California
Anita Swarnakar	San Francisco, California
April J.A. Hafalia	Daly City, California
Shanya D. Becha	San Francisco, California
Narinder K. Chawla	Union City, California
CORRESPONDENCE ADDRESS (Customer No.: 27904)	
INCYTE CORPORATION Legal Department 3160 Porter Drive Palo Alto, California 94304	Phone: (650) 855-0555 Fax: (650) 849-8886 or (650) 845-4166

27904

Atty. Docket No.: PF-1506 P	Type a plus sign (+) inside this box ➡	+

ENCLOSURES

Enclosed are:

1. Return Receipt Postcard;
2. Provisional Application Cover Sheet w/certificate of Express Mailing (2 pp., in duplicate);
3. 103 Pages of Specification (1-103) ;
4. 10 Pages of Claims (104-113) ;
5. 1 Page of Abstract (114) ;
6. 31 Pages of Tables (Tables 1-6) ; and
7. 27 Pages of Sequence Listing (1-27) .

METHOD OF PAYMENT

Applicants hereby authorize the Commissioner to treat any concurrent or future reply in this application that requires a petition for an extension of time under 37 CFR 1.136(a) to be timely, as incorporating a petition for extension of time for the appropriate length of time; and to charge all required fees, including fees under 37 CFR 1.16, 1.17, and all required extension of time fees, or to credit any overpayment, to Deposit Account **09-0108. This sheet is enclosed in duplicate.**

PROVISIONAL FILING FEE AMOUNT: $ 160.00

The invention was not made by an agency of the United States Government or under a contract with an agency of the United States Government.

Respectfully submitted,

INCYTE CORPORATION

Date: 4/30/03

Jenny Buchbinder
Reg. No. 48,588
Direct Dial Telephone: (650) 843-7212

27904

[redacted]

PF-1506 P

KINASES AND PHOSPHATASES

TECHNICAL FIELD

The invention relates to novel nucleic acids, kinases and phosphatases encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and kinases and phosphatases.

BACKGROUND OF THE INVENTION

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups. Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors can activate kinases, which can occur as cell surface receptors or as the activator of the final effector protein, as well as other locations along the signal transduction pathway. Cascades of kinases occur, as well as kinases sensitive to second messenger molecules. This system allows for the amplification of weak signals (low abundance growth factor molecules, for example), as well as the synthesis of many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and cell cycle progression.

KINASES

Kinases comprise the largest known enzyme superfamily and vary widely in their target molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the acceptor molecule, causing internal conformational changes and potentially influencing intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated

proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Book, Vol I, pp. 17-20 Academic Press, San Diego CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

Protein Tyrosine Kinases

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or

nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

One member of the ERK family of MAP kinases, ERK 7, is a novel 61-kDa protein that has motif similarities to ERK1 and ERK2, but is not activated by extracellular stimuli as are ERK1 and ERK2 nor by the common activators, c-Jun N-terminal kinase (JNK) and p38 kinase. ERK7 regulates its nuclear localization and inhibition of growth through its C-terminal tail, not through the kinase

domain as is typical with other MAP kinases (Abe, M.K. (1999) Mol. Cell. Biol. 19:1301-1312).

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) J. Biol. Chem. 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) J. Biol. Chem. 273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms, α, β, γ, δ, and ε. Fish et al. identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al., *supra*).

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state. Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al. have proposed that CKI-epsilon plays a

major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) Science 288:483-491).

Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al. (2000) Biochimie 82:1123-1127). HIPKs contain a conserved protein kinase domain separated from a domain that interacts with homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim, Y.H. et al. (1998) J. Biol. Chem. 273:25875-25879; Wang, Y. et al. (2001) Biochim. Biophys. Acta 1518:168-172). HIPKs act as corepressors for homeodomian transcription factors. This corepressor activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each of which are important in the regulation of cellular protein function (Kim, Y.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:12350-12355).

The human h-warts protein, a homolog of *Drosophila* warts tumor suppressor gene, maps to chromosome 6q24-25.1. It has a serine/threonine kinase domain and is localized to centrosomes in interphase cells. It is involved in mitosis and functions as a component of the mitotic apparatus (Nishiyama, Y. et al. (1999) FEBS Lett. 459:159-165).

Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO J. 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and seratonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be enriched in mammalian forebrain. This protein is associated with vesicles in both axons and dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J.

Neurosci. 14:1-13).

Mitogen-Activated Protein Kinases

The mitogen-activated protein kinases (MAP), which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades, are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) Nature 365:781-783). There are three kinase modules comprising the MAP kinase cascade: MAPK (MAP), MAPK kinase (MAP2K, MAPKK, or MKK), and MKK kinase (MAP3K, MAPKKK, OR MEKK) (Wang,X.S. et al (1998) Biochem. Biophys. Res. Commun. 253:33-37). The extracellular-regulated kinase (ERK) pathway is activated by growth factors and mitogens, for example, epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, or endotoxic lipopolysaccharide (LPS). The closely related though distinct parallel pathways, the c-Jun N-terminal kinase (JNK), or stress-activated kinase (SAPK) pathway, and the p38 kinase pathway are activated by stress stimuli and proinflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.. MAP kinase signaling pathways are present in mammalian cells as well as in yeast.

Cyclin-Dependent Protein Kinases

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) EMBO J. 17:470-481).

Checkpoint and Cell Cycle Kinases

In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which

may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) Science 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) Science 277:1501-1505). Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakarocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-19408). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) J. Biol Chem. 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune diseases, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding

protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) J. Biol. Chem. 273:29066-29071). DRAK1 and DRAK2 are STKs that share homology with the death-associated protein kinases (DAP kinases), known to function in interferon-γ induced apoptosis (Sanjo et al., *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al., *supra*). However, deletion of either the N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al., *supra*).

Mitochondrial Protein Kinases

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) Adv. Enzyme Regul. 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate

dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) *supra*).

KINASES WITH NON-PROTEIN SUBSTRATES

Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) Curr. Opin. Cell. Biol. 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP_2). PIP_2 is then cleaved into inositol triphosphate (IP_3) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) bisphosphate (PIP_2) to PI (3,4,5) P_3 (PIP_3). PIP_3 then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) Biochem. J. 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, *supra*).

An example of lipid kinase phosphorylation activity is the phosphorylation of D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) J. Biol. Chem. 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein

coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., *supra*).

Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) Cancer Res. 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zeleznikar, R.J. et al. (1995) J. Biol. Chem. 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity in order to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenyl cyclase, and production of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and GTP levels also control the activity of certain oncogenic proteins such as $p21^{ras}$ known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) Cancer Res. 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of $p21^{ras}$ and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a

therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buciclovir (Miller, W.H. and R.L. Miller (1980) J. Biol. Chem. 255:7204-7207; Stenberg, K. et al. (1986) J. Biol. Chem. 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and S. Eriksson (1995) Pharmacol. Ther. 67:155-186).

PHOSPHATASES

Protein phosphatases are generally characterized as either serine/threonine- or tyrosine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs, for dual specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues. The protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle and cell signaling processes. Another family of phosphatases is the acid phosphatase or histidine acid phosphatase (HAP) family whose members hydrolyze phosphate esters at acidic pH conditions.

PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal and membranous structures in most tissues, especially the brain. Some PSPs require divalent cations, such as Ca^{2+} or Mn^{2+}, for activity. PSPs play important roles in glycogen metabolism, muscle contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic metabolism (reviewed in Cohen, P. (1989) Annu. Rev. Biochem. 58:453-508). PSPs can be separated into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7. Members of this class are composed of a homologous catalytic subunit bearing a very highly conserved signature sequence, coupled with one or more regulatory subunits (PROSITE PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular localization of PSPs and substrate specificity. The PPM class consists of several closely related isoforms of PP2C and is evolutionarily unrelated to the PPP class.

PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A number of isoforms have been identified, with the alpha and beta forms being produced by alternative splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) Curr. Opin. Neurobiol. 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) Otolaryngol. Head Neck Surg.121:463-468).

PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A et al. (1999) Trends Biosci. 24:186-191). These "B-type" subunits are cell type- and tissue-specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen. Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through the cell cycle (Pallas, D.C. et al. (1992) J. Virol. 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. PP2A, in fact, can dephosphorylate and modulate the activities of more than 30 protein kinases *in vitro*, and other evidence suggests that the same is true *in vivo* for such kinases as PKB, PKC, the calmodulin-dependent kinases, ERK family-MAP kinases, cyclin-dependent kinases, and the IκB kinases (reviewed in Millward et al., *supra*). PP2A is itself a substrate for CKI and CKII kinases, and can be stimulated by polycationic macromolecules. A PP2A-like phosphatase is necessary to maintain the G1 phase destruction of mammalian cyclins A and B (Bastians, H. et al. (1999) Mol. Biol. Cell 10:3927-3941). PP2A is a major activity in the brain and is implicated in regulating neurofilament stability and normal neural function, particularly the phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been proposed to lead to the neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, *supra*).

PP2B, or calcineurin, is a Ca^{2+}-activated dimeric phosphatase and is particularly abundant in the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the calcium/calmodulin complex. Calcineurin is the target of the immunosuppressant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and Mumby, *supra*).

Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci. 22:245-251). One of them, PP5, contains regulatory domains with tetratricopeptide repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids *in vitro* and appears to be involved in a number of signaling pathways, including those controlled by atrial natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent cations (mainly Mn^{2+} or Mg^{2+}) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M. et al. (1998) EMBO J. 17:4744-4752).

In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site consensus sequence includes a cysteine residue which executes a nucleophilic attack on the phosphate moiety during catalysis (Neel, B.G. and N.K. Tonks (1997) Curr. Opin. Cell Biol. 9:193-204). Receptor PTPs are made up of an N-terminal extracellular domain of variable length, a transmembrane region, and a cytoplasmic region that generally contains two copies of the catalytic domain. Although only the first copy seems to have enzymatic activity, the second copy apparently affects the substrate specificity of the first. The extracellular domains of some receptor PTPs contain fibronectin-like repeats, immunoglobulin-like domains, MAM domains (an extracellular motif likely to have an adhesive function), or carbonic anhydrase-like domains (PROSITE PDOC 00323). This wide variety

of structural motifs accounts for the diversity in size and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte activation, and cell proliferation. PTPs μ and κ are involved in cell-cell contacts, perhaps regulating cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility, most of them via the integrin/tyrosine kinase signaling pathway (reviewed in Neel and Tonks, *supra*). CD45 phosphatases regulate signal transduction and lymphocyte activation (Ledbetter, J.A. et al. (1988) Proc. Natl. Acad. Sci. USA 85:8628-8632). Soluble PTPs containing Src-homology-2 domains have been identified (SHPs), suggesting that these molecules might interact with receptor tyrosine kinases. SHP-1 regulates cytokine receptor signaling by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell receptor and c-Kit (reviewed in Neel and Tonks, *supra*). M-phase inducer phosphatase plays a key role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell division (Sadhu, K. et al. (1990) Proc. Natl. Acad. Sci. USA 87:5139-5143). In addition, the genes encoding at least eight PTPs have been mapped to chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma, leukemia, adenocarcinoma, and neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). The PTP enzyme active site comprises the consensus sequence of the MTM1 gene family. The MTM1 gene is responsible for X-linked recessive myotubular myopathy, a congenital muscle disorder that has been linked to Xq28 (Kioschis, P. et al., (1998) Genomics 54:256-266). Many PTKs are encoded by oncogenes, and it is well known that oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may serve to prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine phosphorylation in cells. This is supported by studies showing that overexpression of PTP can suppress transformation in cells and that specific inhibition of PTP can enhance cell transformation (Charbonneau and Tonks, *supra*).

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs. DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C. The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular signal-regulated kinase), JNK (c-Jun N-terminal kinase), and p38 on both tyrosine and threonine residues (PROSITE PDOC 00323, *supra*). In the activated state, these kinases have been implicated in neuronal differentiation, proliferation, oncogenic transformation, platelet aggregation, and apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996) J. Biol. Chem. 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri, D. and M. Ittmann

(1999) Hum. Pathol. 30:419-424) and abnormalities in its expression are associated with numerous cancers (reviewed in Tamura, M. et al. (1999) J. Natl. Cancer Inst. 91:1820-1828).

Histidine acid phosphatase (HAP; EXPASY EC 3.1.3.2), also known as acid phosphatase, hydrolyzes a wide spectrum of substrates including alkyl, aryl, and acyl orthophosphate monoesters and phosphorylated proteins at low pH. HAPs share two regions of conserved sequences, each centered around a histidine residue which is involved in catalytic activity. Members of the HAP family include lysosomal acid phosphatase (LAP) and prostatic acid phosphatase (PAP), both sensitive to inhibition by L-tartrate (PROSITE PDOC00538).

Synaptojanin, a polyphosphoinositide phosphatase, dephosphorylates phosphoinositides at positions 3, 4 and 5 of the inositol ring. Synaptojanin is a major presynaptic protein found at clathrin-coated endocytic intermediates in nerve terminals, and binds the clathrin coat-associated protein, EPS15. This binding is mediated by the C-terminal region of synaptojanin-170, which has 3 Asp-Pro-Phe amino acid repeats. Further, this 3 residue repeat had been found to be the binding site for the EH domains of EPS15 (Haffner, C. et al. (1997) FEBS Lett. 419:175-180). Additionally, synaptojanin may potentially regulate interactions of endocytic proteins with the plasma membrane, and be involved in synaptic vesicle recycling (Brodin, L. et al. (2000) Curr. Opin. Neurobiol. 10:312-320). Studies in mice with a targeted disruption in the synaptojanin 1 gene (Synj1) were shown to support coat formation of endocytic vesicles more effectively than was seen in wild-type mice, suggesting that Synj1 can act as a negative regulator of membrane-coat protein interactions. These findings provide genetic evidence for a crucial role of phosphoinositide metabolism in synaptic vesicle recycling (Cremona, O. et al. (1999) Cell 99:179-188).

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Breast Cancer

More than 180,000 new cases of breast cancer are diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (Gish, K. (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).

Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, *supra*). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

The relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie, K. et al. (1993) Cancer and Metastasis Rev. 12:255-274, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S.S. et al. (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down-regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou, Z. et al. (1998) Int. J. Cancer 78:95-99; Chen, L. et al. (1990) Oncogene 5:1391-1395; Ulrix, W. et al (1999) FEBS Lett 455:23-26; Sager, R. et al. (1996) Curr. Top. Microbiol. Immunol. 213:51-64; and Lee, S.W. et al. (1992) Proc. Natl. Acad. Sci. USA 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer

provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba, I.I. et al. (1998) Clin. Cancer Res. 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

Colon Cancer

While soft tissue sarcomas are relatively rare, more than 50% of new patients diagnosed with the disease will die from it. The molecular pathways leading to the development of sarcomas are relatively unknown, due to the rarity of the disease and variation in pathology. Colon cancer evolves through a multi-step process whereby pre-malignant colonocytes undergo a relatively defined sequence of events leading to tumor formation. Several factors participate in the process of tumor progression and malignant transformation including genetic factors, mutations, and selection.

To understand the nature of gene alterations in colorectal cancer, a number of studies have focused on the inherited syndromes. Familial adenomatous polyposis (FAP), is caused by mutations in the adenomatous polyposis coli gene (APC), resulting in truncated or inactive forms of the protein. This tumor suppressor gene has been mapped to chromosome 5q. Hereditary nonpolyposis colorectal cancer (HNPCC) is caused by mutations in mis-match repair genes. Although hereditary colon cancer syndromes occur in a small percentage of the population and most colorectal cancers are considered sporadic, knowledge from studies of the hereditary syndromes can be generally applied. For instance, somatic mutations in APC occur in at least 80% of sporadic colon tumors. APC mutations are thought to be the initiating event in the disease. Other mutations occur subsequently. Approximately 50% of colorectal cancers contain activating mutations in ras, while 85% contain inactivating mutations in p53. Changes in all of these genes lead to gene expression changes in colon cancer.

Ovarian Cancer

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rate for this disease is very low. Identification of early-stage markers for ovarian cancer would significantly increase the survival rate. Genetic variations involved in ovarian cancer development include mutation of p53 and microsatellite instability. Gene expression patterns likely vary when normal ovary is compared to ovarian tumors.

Alzheimer's Disease

Alzheimer's disease (AD) is a progressive, neurodestructive process of the human neocortex, characterized by the deterioration of memory and higher cognitive function. A progressive and irreversible brain disorder, AD is characterized by three major pathogenic episodes involving (a) an aberrant processing and deposition of beta-amyloid precursor protein (betaAPP) to form neurotoxic

beta-amyloid (betaA) peptides and an aggregated insoluble polymer of betaA that forms the senile plaque, (b) the establishment of intraneuronal neuritic tau pathology yielding widespread deposits of agyrophilic neurofibrillary tangles (NFT) and (c) the initiation and proliferation of a brain-specific inflammatory response. These three seemingly disperse attributes of AD etiopathogenesis are linked by the fact that proinflammatory microglia, reactive astrocytes and their associated cytokines and chemokines are associated with the biology of the microtubule associated protein tau, betaA speciation and aggregation. Missense mutations in the presenilin genes PS1 and PS2, implicated in early onset familial AD, cause abnormal betaAPP processing with resultant overproduction of betaA42 and related neurotoxic peptides. Specific betaA fragments such as betaA42 can further potentiate proinflammatory mechanisms. Expression of the inducible oxidoreductase cyclooxygenase-2 and cytosolic phospholipase A2 (cPLA2) is strongly activated during cerebral ischemia and trauma, epilepsy and AD, indicating the induction of proinflammatory gene pathways as a response to brain injury. Neurotoxic metals such as aluminum and zinc, both implicated in AD etiopathogenesis, and arachidonic acid, a major metabolite of brain cPLA2 activity, each polymerize hyperphosphorylated tau to form NFT-like bundles. Studies have identified a reduced risk for AD in patients aged over 70 years who were previously treated with non-steroidal anti-inflammatory drugs for non-CNS afflictions that include arthritis. (For a review of the interrelationships between the mechanisms of PS1, PS2 and betaAPP gene expression, tau and betaA deposition and the induction, regulation and proliferation in AD of the neuroinflammatory response, see Lukiw, W.J, and Bazan, N.G. (2000) Neurochem. Res. 2000 25:1173-1184).

Lung Cancer

The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of cancer, such as lung cancer. Lung cancer is the leading cause of cancer death in the United States, affecting more than 100,000 men and 50,000 women each year. Nearly 90% of the patients diagnosed with lung cancer are cigarette smokers. Tobacco smoke contains thousands of noxious substances that induce carcinogen metabolizing enzymes and covalent DNA adduct formation in the exposed bronchial epithelium. In nearly 80% of patients diagnosed with lung cancer, metastasis has already occurred. Most commonly lung cancers metastasize to pleura, brain, bone, pericardium, and liver. The decision to treat with surgery, radiation therapy, or chemotherapy is made on the basis of tumor histology, response to growth factors or hormones, and sensitivity to inhibitors or drugs. With current treatments, most patients die within one year of diagnosis. Earlier diagnosis and a systematic approach to identification, staging, and treatment of lung cancer could positively affect patient outcome.

Lung cancers progress through a series of morphologically distinct stages from hyperplasia to invasive carcinoma. Malignant lung cancers are divided into two groups comprising four

histopathological classes. The Non Small Cell Lung Carcinoma (NSCLC) group includes squamous cell carcinomas, adenocarcinomas, and large cell carcinomas and accounts for about 70% of all lung cancer cases. Adenocarcinomas typically arise in the peripheral airways and often form mucin secreting glands. Squamous cell carcinomas typically arise in proximal airways. The histogenesis of squamous cell carcinomas may be related to chronic inflammation and injury to the bronchial epithelium, leading to squamous metaplasia. The Small Cell Lung Carcinoma (SCLC) group accounts for about 20% of lung cancer cases. SCLCs typically arise in proximal airways and exhibit a number of paraneoplastic syndromes including inappropriate production of adrenocorticotropin and anti-diuretic hormone.

Lung cancer cells accumulate numerous genetic lesions, many of which are associated with cytologically visible chromosomal aberrations. The high frequency of chromosomal deletions associated with lung cancer may reflect the role of multiple tumor suppressor loci in the etiology of this disease. Deletion of the short arm of chromosome 3 is found in over 90% of cases and represents one of the earliest genetic lesions leading to lung cancer. Deletions at chromosome arms 9p and 17p are also common. Other frequently observed genetic lesions include overexpression of telomerase, activation of oncogenes such as K-ras and c-myc, and inactivation of tumor suppressor genes such as RB, p53 and CDKN2.

Genes differentially regulated in lung cancer have been identified by a variety of methods. Using mRNA differential display technology, Manda et al. (1999; Genomics 51:5-14) identified five genes differentially expressed in lung cancer cell lines compared to normal bronchial epithelial cells. Among the known genes, pulmonary surfactant apoprotein A and alpha 2 macroglobulin were down regulated whereas nm23H1 was upregulated. Petersen et al.. (2000; Int J. Cancer, 86:512-517) used suppression subtractive hybridization to identify 552 clones differentially expressed in lung tumor derived cell lines, 205 of which represented known genes. Among the known genes, thrombospondin-1, fibronectin, intercellular adhesion molecule 1, and cytokeratins 6 and 18 were previously observed to be differentially expressed in lung cancers. Wang et al. (2000; Oncogene 19:1519-1528) used a combination of microarray analysis and subtractive hybridization to identify 17 genes differentially overexpresssed in squamous cell carcinoma compared with normal lung epithelium. Among the known genes they identified were keratin isoform 6, KOC, SPRC, IGFb2, connexin 26, plakofillin 1 and cytokeratin 13.

Prostate Cancer

As with most tumors, prostate cancer develops through a multistage progression ultimately resulting in an aggressive tumor phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells. Androgen responsive cells become hyperplastic and evolve into early-stage tumors. Although early-stage tumors are often androgen

sensitive and respond to androgen ablation, a population of androgen independent cells evolve from the hyperplastic population. These cells represent a more advanced form of prostate tumor that may become invasive and potentially become metastatic to the bone, brain, or lung. A variety of genes may be differentially expressed during tumor progression. For example, loss of heterozygosity (LOH) is frequently observed on chromosome 8p in prostate cancer. Fluorescence *in situ* hybridization (FISH) revealed a deletion for at least 1 locus on 8p in 29 (69%) tumors, with a significantly higher frequency of the deletion on 8p21.2-p21.1 in advanced prostate cancer than in localized prostate cancer, implying that deletions on 8p22-p21.3 play an important role in tumor differentiation, while 8p21.2-p21.1 deletion plays a role in progression of prostate cancer (Oba, K. et al. (2001) Cancer Genet. Cytogenet. 124: 20-26).

PZ-HPV-7 was derived from epithelial cells cultured from normal tissue from the peripheral zone of the prostate. The cells were transformed by transfection with HPV18. Immunocytochemical analysis showed expression of keratins 5 and 8 and also the early region 6 (E6) oncoprotein of HPV. The cells are negative for prostate specific antigen (PSA).

Interleukin 6 (IL-6) is a multifunctional protein that plays important roles in host defense, acute phase reactions, immune responses, and hematopoiesis. According to the type of biological responses being studied, IL-6 was previously named interferon-b2, 26-kDa protein, B cell stimulatory factor-2 (BSF-2), hybridoma/plasmacytoma growth factor, hepatocyte stimulating factor, cytotoxic T cell differentiation factor, and macrophage-granulocyte inducing factor 2A (MGI-2A). The IL-6 designation was adopted after these variously named proteins were found to be identical on the basis of their amino acid and/or nucleotide sequences. IL-6 is expressed by a variety of normal and transformed cells including T cells, B cells, monocytes/macrophages, fibroblasts, hepatocytes, keratinocytes, astrocytes, vascular endothelial cells, and various tumor cells. The production of IL-6 is upregulated by numerous signals including mitogenic or antigenic stimulation, LPS, calcium ionophore, IL-1, IL-2, IFN, TNF, PDGF, and viruses. IL-4 and IL-13 inhibit IL-6 expression in monocytes.

Obesity

The most important function of adipose tissue is its ability to store and release fat during periods of feeding and fasting. White adipose tissue is the major energy reserve in periods of excess energy use. Its primary purpose is mobilization during energy deprivation. Understanding how various molecules regulate adiposity and energy balance in physiological and pathophysiological situations may lead to the development of novel therapeutics for human obesity. Adipose tissue is also one of the important target tissues for insulin. Adipogenesis and insulin resistance in type II diabetes are linked and present intriguing relations. Most patients with type II diabetes are obese and obesity in turn causes insulin resistance.

NIDDM is the most common form of diabetes mellitus, a chronic metabolic disease that affects

143 million people worldwide. NIDDM is characterized by abnormal glucose and lipid metabolism that results from a combination of peripheral insulin resistance and defective insulin secretion. NIDDM has a complex, progressive etiology and a high degree of heritability. Numerous complications of diabetes including heart disease, stroke, renal failure, retinopathy, and peripheral neuropathy contribute to the high rate of morbidity and mortality.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, kinases and phosphatases, referred to collectively as 'KPP' and individually as 'KPP-1,' 'KPP-2,' 'KPP-3,' 'KPP-4,' 'KPP-5,' 'KPP-6,' 'KPP-7,' 'KPP-8,' 'KPP-9,' 'KPP-10,' 'KPP-11,' 'KPP-12,' 'KPP-13,' and 'KPP-14,' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified kinases and phosphatases and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified kinases and phosphatases and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-14.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID

NO:1-14. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-14. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:15-28.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a polynucleotide comprising a naturally occurring polynucleotide

sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction — amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an

amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence

selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID

NO:15-28, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 6 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and

methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"KPP" refers to the amino acid sequences of substantially purified KPP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of KPP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

An "allelic variant" is an alternative form of the gene encoding KPP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding KPP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as KPP or a polypeptide with at least one functional characteristic of KPP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding KPP, and improper or unexpected hybridization to allelic variants, with a

locus other than the normal chromosomal locus for the polynucleotide encoding KPP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent KPP. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of KPP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of KPP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind KPP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies

which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic"

refers to the capability of the natural, recombinant, or synthetic KPP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding KPP or fragments of KPP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys, Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile

Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of KPP or a polynucleotide encoding KPP which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected

from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:15-28 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:15-28, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:15-28 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:15-28 from related polynucleotides. The precise length of a fragment of SEQ ID NO:15-28 and the region of SEQ ID NO:15-28 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-14 is encoded by a fragment of SEQ ID NO:15-28. A fragment of SEQ ID NO:1-14 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-14. For example, a fragment of SEQ ID NO:1-14 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-14. The precise length of a fragment of SEQ ID NO:1-14 and the region of SEQ ID NO:1-14 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical nucleotide matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in

Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid

sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex

may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of KPP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of KPP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of KPP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of KPP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an KPP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of KPP.

"Probe" refers to nucleic acids encoding KPP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer

binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose

instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing KPP, nucleic acids encoding KPP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic

acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%

or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human kinases and phosphatases (KPP), the polynucleotides encoding KPP, and the use of these compositions for the diagnosis, treatment, or prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows amino acid residues comprising signature sequences, domains, motifs, potential phosphorylation sites, and potential glycosylation sites. Column 5 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are kinases and phosphatases. For example, SEQ ID NO:11 is 78% identical, from residue M1 to residue W1219, to mouse NIK (GenBank ID g1872546)

as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:11 also has homology to proteins that activate the c-Jun N-terminal kinase (Mapk8) signaling pathway, and are mitogen-activated protein kinase kinase kinase kinases (MAP4K), as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:11 also contains a CNH domain, a protein kinase domain, a domain found in NIK1-like kinases, and a serine/threonine kinase catalytic domain, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:11 is a protein kinase. SEQ ID NO:1-10 and SEQ ID NO:12-14 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-14 are described in Table 5.

As shown in Table 4, full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 lists fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:15-28 or that distinguish between SEQ ID NO:15-28 and related polynucleotides. Column 3 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide embodiments. Columns 4 and 5 of Table 4 show the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences in column 3 relative to their respective full length sequences.

The identification numbers in Column 3 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 9004061 5J1). Alternatively, the identification numbers in column 3 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the identification numbers in column 3 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 3 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records

(*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 3 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 3 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is the identification number of a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 6 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of

the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses KPP variants. Various embodiments of KPP variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the KPP amino acid sequence, and can contain at least one functional or structural characteristic of KPP.

Various embodiments also encompass polynucleotides which encode KPP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:15-28, which encodes KPP. The polynucleotide sequences of SEQ ID NO:15-28, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding KPP. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding KPP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:15-28 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:15-28. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of KPP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding KPP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding KPP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing during mRNA processing. A splice variant may have less than about 70%, or alternatively less than

about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding KPP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding KPP. For example, a polynucleotide comprising a sequence of SEQ ID NO:19, a polynucleotide comprising a sequence of SEQ ID NO:20 and a polynucleotide comprising a sequence of SEQ ID NO:21 are splice variants of each other. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of KPP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding KPP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring KPP, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode KPP and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring KPP under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding KPP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding KPP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode KPP and KPP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding KPP or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:15-28 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding KPP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include

sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode KPP may be cloned in recombinant DNA molecules that direct expression of KPP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express KPP.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter KPP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of KPP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random

point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding KPP may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, KPP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of KPP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active KPP, the polynucleotides encoding KPP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding KPP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding KPP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding KPP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl.

Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding KPP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding KPP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding KPP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding KPP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding KPP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of KPP are needed, e.g. for the production of antibodies, vectors which direct high level expression of KPP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may

be used.

Yeast expression systems may be used for production of KPP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of KPP. Transcription of polynucleotides encoding KPP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding KPP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses KPP in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of KPP in cell lines is preferred. For example, polynucleotides encoding KPP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a

selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding KPP is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding KPP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding KPP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding KPP and that express KPP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of KPP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on KPP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990)

Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding KPP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding KPP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding KPP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode KPP may be designed to contain signal sequences which direct secretion of KPP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding KPP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric KPP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of KPP activity. Heterologous protein and peptide moieties

may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the KPP encoding sequence and the heterologous protein sequence, so that KPP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled KPP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

KPP, fragments of KPP, or variants of KPP may be used to screen for compounds that specifically bind to KPP. One or more test compounds may be screened for specific binding to KPP. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to KPP. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of KPP can be used to screen for binding of test compounds, such as antibodies, to KPP, a variant of KPP, or a combination of KPP and/or one or more variants KPP. In an embodiment, a variant of KPP can be used to screen for compounds that bind to a variant of KPP, but not to KPP having the exact sequence of a sequence of SEQ ID NO:1-14. KPP variants used to perform such screening can have a range of about 50% to about 99% sequence identity to KPP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to KPP can be closely related to the natural ligand of KPP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor KPP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to KPP can be closely related to the natural receptor to which KPP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for KPP which is capable of propagating a signal, or a decoy receptor for KPP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to KPP, fragments of KPP, or variants of KPP. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of KPP. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of KPP. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of KPP.

In an embodiment, anticalins can be screened for specific binding to KPP, fragments of KPP, or variants of KPP. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit KPP involves producing appropriate cells which express KPP, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing KPP or cell membrane fractions which contain KPP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either KPP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with KPP, either in solution or affixed to a solid support, and detecting the binding of KPP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

KPP, fragments of KPP, or variants of KPP may be used to screen for compounds that modulate the activity of KPP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for KPP activity, wherein KPP is combined with at least one test compound, and the activity of KPP in the presence of a test compound is compared with the activity of KPP in the absence of the test compound. A change in the activity of KPP in the presence of the test compound is indicative of a compound that modulates the activity of KPP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising KPP under conditions suitable for KPP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of KPP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding KPP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted

by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding KPP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding KPP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding KPP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress KPP, e.g., by secreting KPP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of KPP and kinases and phosphatases. In addition, examples of tissues expressing KPP can be found in Example XI. Therefore, KPP appears to play a role in cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. In the treatment of disorders associated with increased KPP expression or activity, it is desirable to decrease the expression or activity of KPP. In the treatment of disorders associated with decreased KPP expression or activity, it is desirable to increase the expression or activity of KPP.

Therefore, in one embodiment, KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP. Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial

dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary

ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid-lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue

disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease.

In another embodiment, a vector capable of expressing KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified KPP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of KPP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those listed above.

In a further embodiment, an antagonist of KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP. Examples of such disorders include, but are not limited to, those cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers described above. In one aspect, an antibody which specifically binds KPP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express KPP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of KPP may be produced using methods which are generally known in the art.

In particular, purified KPP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind KPP. Antibodies to KPP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e., those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with KPP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to KPP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of KPP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to KPP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce KPP-specific single chain antibodies.

Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for KPP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between KPP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering KPP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for KPP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of KPP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple KPP epitopes, represents the average affinity, or avidity, of the antibodies for KPP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular KPP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the KPP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of KPP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine

the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of KPP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding KPP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding KPP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding KPP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) FASEB J. 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271-278; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding KPP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated

cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in KPP expression or regulation causes disease, the expression of KPP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in KPP are treated by constructing mammalian expression vectors encoding KPP and introducing these vectors by mechanical means into KPP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of KPP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). KPP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding KPP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to KPP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding KPP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding KPP to cells which have one or more genetic abnormalities with respect to the expression of KPP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding KPP to target cells which have one or more genetic abnormalities with respect to the expression of KPP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing KPP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a

reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding KPP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for KPP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of KPP-coding RNAs and the synthesis of high levels of KPP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of KPP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have

been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding KPP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding KPP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence

complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. siRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. siRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

siRNA can be generated indirectly by introduction of dsRNA into the targeted cell. Alternatively, siRNA can be synthesized directly and introduced into a cell by transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable siRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected siRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue,

shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene can be determined, for example, by northern analysis methods using the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined, for example, by microarray methods; by polyacrylamide gel electrophoresis; and by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding KPP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased KPP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding KPP may be therapeutically useful, and in the treatment of disorders associated with decreased KPP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding KPP may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding KPP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding KPP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding KPP. The amount of hybridization may be quantified, thus forming the

basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of KPP, antibodies to KPP, and mimetics, agonists, antagonists, or inhibitors of KPP.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled

the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising KPP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, KPP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example KPP or fragments thereof, antibodies of KPP, and agonists, antagonists or inhibitors of KPP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy.

Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind KPP may be used for the diagnosis of disorders characterized by expression of KPP, or in assays to monitor patients being treated with KPP or agonists, antagonists, or inhibitors of KPP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for KPP include methods which utilize the antibody and a label to detect KPP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring KPP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of KPP expression. Normal or standard values for KPP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to KPP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of KPP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding KPP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of KPP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of KPP, and to monitor regulation of KPP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding KPP or closely related molecules may be used to identify nucleic acid sequences which encode KPP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved

motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding KPP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the KPP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:15-28 or from genomic sequences including promoters, enhancers, and introns of the KPP gene.

Means for producing specific hybridization probes for polynucleotides encoding KPP include the cloning of polynucleotides encoding KPP or KPP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding KPP may be used for the diagnosis of disorders associated with expression of KPP. Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder

such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental

retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease. Polynucleotides encoding KPP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered KPP expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding KPP may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding KPP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding KPP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the

treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of KPP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding KPP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding KPP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding KPP, or a fragment of a polynucleotide complementary to the polynucleotide encoding KPP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding KPP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding KPP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs

in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of KPP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes

simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, KPP, fragments of KPP, or antibodies specific for KPP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality

signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may

be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for KPP to quantify the levels of KPP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/25116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding KPP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding KPP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, KPP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes

between KPP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with KPP, or fragments thereof, and washed. Bound KPP is then detected by methods well known in the art. Purified KPP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding KPP specifically compete with a test compound for binding KPP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with KPP.

In additional embodiments, the nucleotide sequences which encode KPP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs are derived from cDNA libraries described in the LIFESEQ database (Incyte, Palo Alto CA) and shown in Table 4, column 3. Some tissues are homogenized and lysed in guanidinium isothiocyanate, while others are homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates are centrifuged over CsCl cushions or extracted with chloroform. RNA is precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA are repeated as necessary to increase RNA purity. In some cases, RNA is treated with DNase. For most libraries, poly(A)+ RNA is isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth

CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA is isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene is provided with RNA and constructs the corresponding cDNA libraries. Otherwise, cDNA is synthesized and cDNA libraries are constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription is initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters are ligated to double stranded cDNA, and the cDNA is digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA is size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs are ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte, Palo Alto CA), pRARE (Incyte), or pINCY (Incyte), or derivatives thereof. Recombinant plasmids are transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I are recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids are purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids are resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA is amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps are carried out in a single reaction mixture. Samples are processed and stored in 384-well plates, and the concentration of amplified plasmid DNA is quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II are sequenced as follows. Sequencing reactions are processed using standard methods or high-throughput instrumentation such

as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions are prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides are carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences are identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences are selected for extension using the techniques disclosed in Example VIII.

Polynucleotide sequences derived from Incyte cDNAs are validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof are then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries are performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences are assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) are used to extend Incyte cDNA assemblages to full length. Assembly is performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages are screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences are translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences are subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as

PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 5 summarizes tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences are also used to identify polynucleotide sequence fragments from SEQ ID NO:15-28. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative kinases and phosphatases are initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once is set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode kinases and phosphatases, the encoded polypeptides are analyzed by querying against PFAM models for kinases and phosphatases. Potential kinases and phosphatases are also identified by homology to Incyte cDNA sequences that have been annotated as kinases and phosphatases. These selected Genscan-predicted sequences are then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences are then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis is also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA

coverage is available, this information is used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences are obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences are derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences are extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III are mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster is analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that are subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval is present on more than one sequence in the cluster are identified, and intervals thus identified are considered to be equivalent by transitivity. For example, if an interval is present on a cDNA and two genomic sequences, then all three intervals are considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified are then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) are given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences are translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan are corrected by comparison to the top BLAST hit from genpept. Sequences are further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences are extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III are queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog is then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein is generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog,

the chimeric protein, or both are used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences are therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences are examined to determine whether they contain a complete gene.

VI. Chromosomal Mapping of KPP Encoding Polynucleotides

The sequences used to assemble SEQ ID NO:15-28 are compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:15-28 are assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon are used to determine if any of the clustered sequences have been previously mapped. Inclusion of a mapped sequence in a cluster results in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST are used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding KPP are analyzed with respect to the tissue sources from which they are derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding KPP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ database (Incyte, Palo Alto CA).

VIII. Extension of KPP Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer is synthesized to initiate 5' extension of the known fragment, and the other primer is synthesized to initiate 3' extension of the known fragment. The initial primers are designed using OLIGO 4.06 software (National

Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations is avoided.

Selected human cDNA libraries are used to extend the sequence. If more than one extension is necessary or desired, additional or nested sets of primers are designed.

High fidelity amplification is obtained by PCR using methods well known in the art. PCR is performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contains DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ are as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well is determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate is scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture is analyzed by electrophoresis on a 1 % agarose gel to determine which reactions are successful in extending the sequence.

The extended nucleotides are desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides are separated on low concentration (0.6 to 0.8%) agarose gels, fragments are excised, and agar digested with Agar ACE (Promega). Extended clones are religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells are selected on antibiotic-containing media, and individual colonies are picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells are lysed, and DNA is amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated

29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA is quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries are reamplified using the same conditions as described above. Samples are diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in KPP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) are identified in SEQ ID NO:15-28 using the LIFESEQ database (Incyte). Sequences from the same gene are clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters is used to distinguish SNPs from other sequence variants. Preliminary filters remove the majority of basecall errors by requiring a minimum Phred quality score of 15, and remove sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis is applied to the original chromatogram files in the vicinity of the putative SNP. Clone error filters use statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters use statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removes duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs are selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprises 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprises 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprises 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprises 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies are first analyzed in the Caucasian population; in some cases those SNPs which show no allelic variance in this population are not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:15-28 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a

fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma-Aldrich, St. Louis MO) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic

apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different

fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be differentially expressed.

Expression

For example, expression of SEQ ID NO:16, SEQ ID NO:22, SEQ ID NO:24-25, and SEQ ID NO:28 were differentially expressed in breast carcinoma cell lines versus a cell line derived from normal breast epithelial tissue as determined by microarray analysis. Gene expression profiles of nonmalignant mammary epithelial cells were compared to gene expression profiles of various breast carcinoma lines at different stages of tumor progression. The cells were grown in defined serum-free H14 medium to 70-80% confluence prior to RNA harvest. Cell lines compared included: a) HMEC, a primary breast epithelial cell line isolated from a normal donor, b) MCF-10A, a breast mammary gland cell line isolated from a 36-year-old woman with fibrocystic breast disease, c) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, d) T-47D, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with an infiltrating ductal carcinoma of the breast, e) Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female, f) BT-20, a breast carcinoma cell line derived *in vitro* from cells emigrating out of thin slices of the tumor mass isolated from a 74-year-old female, g) MDA-mb-231, a breast tumor cell line isolated from the pleural effusion of a 51-year-old female, and h) MDA-mb-435S, a spindle-shaped strain that evolved from the parent line (435) isolated by R. Cailleau from pleural effusion of a 31-year-old female with metastatic, ductal adenocarcinoma of the breast. Expression of SEQ ID NO:24 was increased at least two-fold in MCF7 cells versus HMECs. In a similar experiment, expression of SEQ ID NO:22 was decreased at least two-fold in Sk-BR-3 cells

versus HMECs. In a similar experiment, expression of SEQ ID NO:25 was decreased at least two-fold in Sk-BR-3, T-47D, and MCF7 cells versus HMECs. In a similar experiment, expression of SEQ ID NO:28 was decreased at least two-fold in MDA-mb-231 and MCF-10A cells versus HMECs. Therefore, in various embodiments, SEQ ID NO:16, SEQ ID NO:22, SEQ ID NO:24-25, and SEQ ID NO:28 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

In another example, expression of SEQ ID NO:16, SEQ ID NO:22, SEQ ID NO:24-25, and SEQ ID NO:28 were differentially expressed in breast carcinoma cell lines versus a cell line derived from a donor with non-malignant, fibrocystic breast disease as determined by microarray analysis. Gene expression profiles of nonmalignant mammary epithelial cells were compared to gene expression profiles of various breast carcinoma lines at different stages of tumor progression. The cells were grown in defined serum-free TCH medium, defined serum-free H14 medium, or the supplier's recommended medium to 70-80% confluence prior to RNA harvest and compared to MCF-10A cells grown in the same medium. Cell lines compared included: a) MCF-10A, a breast mammary gland (luminal ductal characteristics) cell line isolated from a 36-year-old woman with fibrocystic breast disease; b) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, c) T-47D, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with an infiltrating ductal carcinoma of the breast, d) Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female, e) BT-20, a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of the tumor mass isolated from a 74-year-old female, f) MDA-mb-231, a breast tumor cell line isolated from the pleural effusion of a 51-year old female, and g) MDA-mb-435S, a spindle shaped strain that evolved from the parent line (435) isolated from the pleural effusion of a 31-year-old female with metastatic, ductal adenocarcinoma of the breast. . Expression of SEQ ID NO:16 was increased at least two-fold in MCF7 cells when grown in either the defined serum-free H14 medium or the supplier's recommended medium as compared with MCF-10A cells grown under the same conditions. In a similar experiment, expression of SEQ ID NO:22 was decreased at least two-fold in Sk-BR-3 cells when grown in any of the growth conditions as compared with MCF-10A cells grown under the same conditions. In a similar experiment, expression of SEQ ID NO:24 was increased at least two-fold in MCF7 cells when grown in any of the growth conditions as compared with MCF-10A cells grown under the same conditions. In a similar experiment, expression of SEQ ID NO:25 was decreased at least two-fold in Sk-BR-3 cells and T-47D cells when grown in any of the growth conditions as compared with MCF-10A cells grown under the same conditions. In a similar experiment, expression of SEQ ID NO:28 was increased at least two-fold in MDA-mb-231 cells when grown in either the defined serum-free H14 medium or the

supplier's recommended medium as compared with MCF-10A cells grown under the same conditions. Therefore, in various embodiments, SEQ ID NO:16, SEQ ID NO:22, SEQ ID NO:24-25, and SEQ ID NO:28 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

In another example, expression of SEQ ID NO:18 was down-regulated in a breast cancer cell line (MCF7) treated with TNFα versus untreated MCF7 cells as determined by microarray analysis. MCF7 cells were treated with 10 ng/mL TNFα for 1, 4, 8, 12, 24, 48, and 72 hours. Treated cells were compared to untreated cells kept in culture for the same amount of time. Expression of SEQ ID NO:18 was decreased at least two-fold in MCF7 cells treated with 10 ng/mL TNFα for 4, 8, 24, or 48 hours as compared with untreated MCF7 cells. Therefore, in various embodiments, SEQ ID NO:18 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

In another example, expression of SEQ ID NO:22 was down-regulated in ovary tumor tissue versus normal ovary tissue as determined by microarray analysis. Expression of SEQ ID NO:22 was decreased at least two-fold in ovary tumor tissue as compared with matched normal ovary tissue from the same donor in 1 of 2 donors tested. Therefore, in various embodiments, SEQ ID NO:22 can be used for one or more of the following: i) monitoring treatment of ovarian cancer, ii) diagnostic assays for ovarian cancer, and iii) developing therapeutics and/or other treatments for ovarian cancer.

For example, expression of SEQ ID NO:25 was down-regulated in brain tissue from donors with Alzheimer's disease (AD) versus brain tissue from a normal donor as determined by microarray analysis. Specific dissected brain regions from the cerebellum, dentate nucleus, and vermis of a normal donor were compared to: a) the corresponding regions dissected from the brain of a female with mild AD; b) the corresponding regions dissected from the brain of a female with severe AD. The diagnosis of normal or mild AD was established by a certified neuropathologist based on microscopic examination of multiple sections throughout the brain. Expression of SEQ ID NO:25 was decreased at least two-fold in the striatum and globus pallidus region of the brain of a donor with severe AD and a donor with mild AD as compared with the corresponding region of the brain from a normal donor. Therefore, in various embodiments, SEQ ID NO:25 can be used for one or more of the following: i) monitoring treatment of AD, ii) diagnostic assays for AD, and iii) developing therapeutics and/or other treatments for AD.

In another example, expression of SEQ ID NO:28 was up-regulated in lung tumor tissue versus normal lung tissue as determined by microarray analysis. Expression of SEQ ID NO:28 was increased at least two-fold in lung tumor tissue as compared with matched normal lung tissue from the same donor in 3 of 4 donors tested. Therefore, in various embodiments, SEQ ID NO:28 can be used for one

or more of the following: i) monitoring treatment of lung cancer, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

In another example, expression of SEQ ID NO:28 was down-regulated to a lesser extent in preadipocytes taken from an obese donor versus preadipocytes taken from a non-obese donor as determined by microarray analysis. Primary subcutaneous preadipocytes were isolated from the adipose tissue of a non-obese donor, a 28-year-old healthy female with body mass index (BMI) of 23.59, and an obese donor, a 40-year-old healthy female with a body mass index (BMI) of 32.47. The preadipocytes from each donor were cultured and induced to differentiate into adipocytes by growing them in differentiation medium containing PPAR-γ agonist and human insulin (Zen-Bio). Some thiazolidinediones or PPAR-γ agonists, which bind and activate an orphan nuclear receptor, PPAR-γ, have been shown to induce human adipocyte differentiation. The preadipocytes were treated with human insulin and PPAR-γ agonist for 3 days and subsequently were switched to medium containing insulin for a range of time periods ranging from one to 20 days before the cells were collected for analysis. Differentiated adipocytes from each donor were compared to untreated preadipocytes, maintained in culture in the absence of differentiation-inducing agents, from the same donor. Between 80% and 90% of the preadipocytes finally differentiated to adipocytes as observed under phase contrast microscopy. Expression of SEQ ID NO:28 was decreased at least two-fold in differentiated preadipocytes from a normal donor versus non-differentiated preadipocytes from the same donor. In contrast, no differential expression was seen in differentiated preadipocytes from an obese donor versus non-differentiated preadipocytes from the same donor. These data suggest that SEQ ID NO:28 is differentially expressed in adipocytes from normal subjects but not in adipocytes from obese subjects. Therefore, in various embodiments, SEQ ID NO:28 can be used for one or more of the following: i) monitoring treatment of diabetes mellitus and other disorders, such as obesity and hypertension ii) diagnostic assays for diabetes mellitus and other disorders, such as obesity and hypertension iii) developing therapeutics and/or other treatments for diabetes mellitus and other disorders, such as obesity and hypertension.

In another example, SEQ ID NO:18, SEQ ID NO:25, and SEQ ID NO:27 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to

another. The expression of SEQ ID NO:18 was increased by at least two-fold in small intestine and liver as compared to the reference sample. Therefore, SEQ ID NO:18 can be used as a tissue marker for small intestine and liver. The expression of SEQ ID NO:25 was increased by at least two-fold in brain (temporal cortex) and leukocytes as compared to the reference sample. Therefore, SEQ ID NO:25 can be used as a tissue marker for brain (temporal cortex) and leukocytes. The expression of SEQ ID NO:27 was increased by at least two-fold in brain as compared to the reference sample. Therefore, SEQ ID NO:27 can be used as a tissue marker for brain.

For example, SEQ ID NO:15 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:15 was increased by at least two-fold in leukocytes, thymus gland, and tonsil as compared to the reference sample. Therefore, SEQ ID NO:15 can be used as a tissue marker for leukocytes, thymus gland, and tonsil.

For example, SEQ ID NO:19-21 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:19-21 was increased by at least two-fold in muscle, adipose tissue, and liver as compared to the reference sample. Therefore, SEQ ID NO 19-21 can be used as a tissue marker for muscle, adipose tissue, and liver.

XII. Complementary Polynucleotides

Sequences complementary to the KPP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring KPP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of KPP. To inhibit transcription, a

complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the KPP-encoding transcript.

XIII. Expression of KPP

Expression and purification of KPP is achieved using bacterial or virus-based expression systems. For expression of KPP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express KPP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of KPP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding KPP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, KPP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from KPP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified KPP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, XIX, XX, and XXI, where applicable.

XIV. Functional Assays

KPP function is assessed by expressing the sequences encoding KPP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector

containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of KPP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding KPP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding KPP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of KPP Specific Antibodies

KPP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the KPP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well

described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-KPP activity by, for example, binding the peptide or KPP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring KPP Using Specific Antibodies

Naturally occurring or recombinant KPP is substantially purified by immunoaffinity chromatography using antibodies specific for KPP. An immunoaffinity column is constructed by covalently coupling anti-KPP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing KPP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of KPP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/KPP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and KPP is collected.

XVII. Identification of Molecules Which Interact with KPP

KPP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled KPP, washed, and any wells with labeled KPP complex are assayed. Data obtained using different concentrations of KPP are used to calculate values for the number, affinity, and association of KPP with the candidate molecules.

Alternatively, molecules interacting with KPP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

KPP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of KPP Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein substrate by KPP in the presence of [γ-^{32}P]ATP. KPP is incubated with the protein substrate, ^{32}P-ATP,

and an appropriate kinase buffer. The ^{32}P incorporated into the substrate is separated from free ^{32}P-ATP by electrophoresis and the incorporated ^{32}P is counted using a radioisotope counter. The amount of incorporated ^{32}P is proportional to the activity of KPP. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma ^{32}P-ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated ^{32}P-peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma ^{32}P-ATP. The reservoir of the centrifuged unit containing the ^{32}P-peptide product as retentate is then counted in a scintillation counter. This procedure allows the assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes include but are not limited to: Histone H1 (Sigma) and $p34^{cdc2}$kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) Methods Enzymol. 200:62-81).

In another alternative, protein kinase activity of KPP is demonstrated in an assay containing KPP, 50 µl of kinase buffer, 1 µg substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10 µg ATP, and 0.5 µCi [γ-^{32}P]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [γ-^{32}P]ATP is removed by washing and the incorporated radioactivity is measured using a scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and resolved on a 12% SDS polyacrylamide gel followed by autoradiography. The amount of incorporated ^{32}P is proportional to the activity of KPP.

In yet another alternative, adenylate kinase or guanylate kinase activity of KPP may be measured by the incorporation of ^{32}P from [γ-^{32}P]ATP into ADP or GDP using a gamma radioisotope counter. KPP, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and ^{32}P-labeled ATP as the phosphate donor. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is excised and counted. The radioactivity recovered is proportional to the activity of KPP.

In yet another alternative, other assays for KPP include scintillation proximity assays (SPA),

scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of KPP activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

In another alternative, phosphatase activity of KPP is measured by the hydrolysis of para-nitrophenyl phosphate (PNPP). KPP is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62). Alternatively, acid phosphatase activity of KPP is demonstrated by incubating KPP-containing extract with 100 μl of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 μl of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of KPP in the assay.

In the alternative, KPP activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM KPP in a final volume of 30 μl containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% β-mercaptoethanol and 10 μM substrate, ^{32}P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched with 450 μl of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $Na_4P_2O_7$, and 2 mM NaH_2PO_4, then centrifuged at 12,000 × *g* for 5 min. Acid-soluble ^{32}Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

XIX. Kinase Binding Assay

Binding of KPP to a FLAG-CD44 cyt fusion protein can be determined by incubating KPP with anti-KPP-conjugated immunoaffinity beads followed by incubating portions of the beads (having 10-20 ng of protein) with 0.5 ml of a binding buffer (20 mM Tris-HCL (pH 7.4), 150 mM NaCl, 0.1% bovine serum albumin, and 0.05% Triton X-100) in the presence of ^{125}I-labeled FLAG-CD44cyt fusion protein (5,000 cpm/ng protein) at 4 °C for 5 hours. Following binding, beads were washed thoroughly in the binding buffer and the bead-bound radioactivity measured in a scintillation counter (Bourguignon, L.Y.W. et al. (2001) J. Biol. Chem. 276:7327-7336). The amount of incorporated ^{32}P is proportional to the amount of bound KPP.

XX. Identification of KPP Inhibitors

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. KPP activity is measured for each well and the ability of each compound to inhibit KPP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules

which enhance KPP activity.

XXI. Identification of KPP Substrates

A KPP "substrate-trapping" assay takes advantage of the increased substrate affinity that may be conferred by certain mutations in the PTP signature sequence of protein tyrosine phosphatases. KPP bearing these mutations form a stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a clone encoding the catalytic domain of KPP is performed using a method standard in the art or a commercial kit, such as the MUTA-GENE kit from BIO-RAD. For expression of KPP mutants in *Escherichia coli*, DNA fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding KPP or a glutathione S-transferase (GST)-KPP fusion protein. KPP mutants are expressed in *E. coli* and purified by chromatography.

The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated transfection with 20 μg of CsCl-purified DNA per 10-cm dish of cells or 8 μg per 6-cm dish. Forty-eight hours after transfection, cells are stimulated with 100 ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are lysed in 50 mM Tris·HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 μg/ml leupeptin/5 μg/ml aprotinin/1 mM benzamidine (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). KPP is immunoprecipitated from lysates with an appropriate antibody. GST-KPP fusion proteins are precipitated with glutathione-Sepharose, 4 μg of mAb or 10 μl of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or further purified to identify substrate molecules (Flint, A.J. et al. (1997) Proc. Natl. Acad. Sci. USA 94:1680-1685).

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

PF-1506 P

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1,

c) a polypeptide consisting essentially of a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2-3, SEQ ID NO:8, SEQ ID NO:10, and SEQ ID NO:12-14,

d) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:4 and SEQ ID NO:11,

e) a polypeptide comprising a naturally occurring amino acid sequence at least 91% identical to the amino acid sequence of SEQ ID NO:5,

f) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:6-7,

g) a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to the amino acid sequence of SEQ ID NO:9,

h) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and

i) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28,

b) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-22 and SEQ ID NO:26-28,

c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 96% identical to the polynucleotide sequence of SEQ ID NO:23,

d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 98% identical to the polynucleotide sequence of SEQ ID NO:24,

e) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 95% identical to the polynucleotide sequence of SEQ ID NO:25,

f) a polynucleotide complementary to a polynucleotide of a),

g) a polynucleotide complementary to a polynucleotide of b),

h) a polynucleotide complementary to a polynucleotide of c),

i) a polynucleotide complementary to a polynucleotide of d),

j) a polynucleotide complementary to a polynucleotide of e), and

k) an RNA equivalent of a)-j).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

19. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional KPP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test

compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A method for a diagnostic test for a condition or disease associated with the expression of KPP in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the

presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, further comprising a label.

35. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from the animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14 in a sample, the method comprising:

a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14 from a sample, the method comprising:

a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID

NO:1-14.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

ABSTRACT OF THE DISCLOSURE

Various embodiments of the invention provide human kinases and phosphatases (KPP) and polynucleotides which identify and encode KPP. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of KPP.

PF-1506 P

<110> ELLIOTT, Vicki S.; KHARE, Reena;
RICHARDSON, Thomas W.; MARQUIS, Joseph P.;
SWARNAKAR, Anita; HAFALIA, April J.A.;
BECHA, Shanya D.; CHAWLA, Narinder K.

<120> KINASES AND PHOSPHATASES

<130> PF-1506 P

<140> To Be Assigned
<141> Herewith

<160> 28
<170> PERL Program

<210> 1
<211> 83
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7517831CD1

<400> 1
Met Gly Cys Gly Cys Ser Ser His Pro Glu Asp Asp Trp Met Glu
1 5 10 15
Asn Ile Asp Val Cys Glu Asn Cys His Tyr Pro Ile Val Pro Leu
20 25 30
Asp Gly Lys Gly Thr Leu Leu Ile Arg Asn Gly Ser Glu Thr Thr
35 40 45
Trp Leu Ser Leu Cys Thr Ala Met Ser Pro Leu Thr Thr Glu Ile
50 55 60
Trp Ala Leu Arg Arg Gly Asn Ser Ser Ala Ser Trp Ser Arg Ala
65 70 75
Ala Ser Gly Gly Arg Arg Ser Pro
80

<210> 2
<211> 292
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7520272CD1

<400> 2
Met Ala Asp Gln Ala Pro Phe Asp Thr Asp Val Asn Thr Leu Thr
1 5 10 15
Arg Phe Val Met Glu Glu Gly Arg Lys Ala Arg Gly Thr Gly Glu
20 25 30
Leu Thr Gln Leu Leu Asn Ser Leu Cys Thr Ala Val Lys Ala Ile
35 40 45
Ser Ser Ala Val Arg Lys Ala Gly Ile Ala His Leu Tyr Gly Ile
50 55 60
Ala Gly Ser Thr Asn Val Thr Gly Asp Gln Val Lys Lys Leu Asp
65 70 75
Val Leu Ser Asn Asp Leu Val Met Asn Met Leu Lys Ser Ser Phe
80 85 90
Ala Thr Cys Val Leu Val Ser Glu Glu Asp Lys His Ala Ile Ile
95 100 105
Val Glu Pro Glu Lys Arg Gly Lys Tyr Val Val Cys Phe Asp Pro
110 115 120
Leu Asp Gly Ser Ser Asn Ile Asp Cys Leu Val Ser Val Gly Thr
125 130 135

PF-1506 P

```
Ile Phe Gly Ile Tyr Arg Lys Lys Ser Thr Asp Glu Pro Ser Glu
                140                 145                 150
Lys Asp Ala Leu Gln Pro Gly Arg Asn Leu Val Ala Ala Gly Tyr
                155                 160                 165
Ala Leu Tyr Gly Ser Ala Thr Met Leu Val Leu Ala Met Asp Cys
                170                 175                 180
Gly Val Asn Cys Phe Met Leu Asp Pro Asp Asn Ser Ala Pro Tyr
                185                 190                 195
Gly Ala Arg Tyr Val Gly Ser Met Val Ala Asp Val His Arg Thr
                200                 205                 210
Leu Val Tyr Gly Gly Ile Phe Leu Tyr Pro Ala Asn Lys Lys Ser
                215                 220                 225
Pro Asn Gly Lys Leu Arg Leu Leu Tyr Glu Cys Asn Pro Met Ala
                230                 235                 240
Tyr Val Met Glu Lys Ala Gly Gly Met Ala Thr Thr Gly Lys Glu
                245                 250                 255
Ala Val Leu Asp Val Ile Pro Thr Asp Ile His Gln Arg Ala Pro
                260                 265                 270
Val Ile Leu Gly Ser Pro Asp Asp Val Leu Glu Phe Leu Lys Val
                275                 280                 285
Tyr Glu Lys His Ser Ala Gln
                290
```

<210> 3
<211> 434
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7521279CD1

<400> 3

```
Met Ala Ser Pro Arg Glu Leu Thr Gln Asn Pro Leu Lys Lys Ile
  1                 5                  10                  15
Trp Met Pro Tyr Ser Asn Gly Arg Pro Ala Leu His Ala Cys Gln
                 20                  25                  30
Arg Gly Val Cys Met Thr Asn Cys Pro Thr Leu Ile Val Met Val
                 35                  40                  45
Gly Leu Pro Ala Arg Gly Lys Thr Tyr Ile Ser Lys Lys Leu Thr
                 50                  55                  60
Arg Tyr Leu Asn Trp Ile Gly Val Pro Thr Arg Glu Phe Asn Val
                 65                  70                  75
Gly Gln Tyr Arg Arg Asp Val Val Lys Thr Tyr Lys Ser Phe Glu
                 80                  85                  90
Phe Phe Leu Pro Asp Asn Glu Glu Gly Leu Lys Ile Arg Lys Gln
                 95                 100                 105
Cys Ala Leu Ala Ala Leu Arg Asp Val Arg Arg Phe Leu Ser Glu
                110                 115                 120
Glu Gly Gly His Val Ala Val Phe Asp Ala Thr Asn Thr Thr Arg
                125                 130                 135
Glu Arg Arg Ala Thr Ile Phe Asn Phe Gly Glu Gln Asn Gly Tyr
                140                 145                 150
Lys Thr Phe Phe Val Glu Ser Ile Cys Val Asp Pro Glu Val Ile
                155                 160                 165
Ala Ala Asn Ile Val Gln Val Lys Leu Gly Ser Pro Asp Tyr Val
                170                 175                 180
Asn Arg Asp Ser Asp Glu Ala Thr Glu Asp Phe Met Arg Arg Ile
                185                 190                 195
Glu Cys Tyr Glu Asn Ser Tyr Glu Ser Leu Asp Glu Asp Leu Asp
                200                 205                 210
Arg Asp Leu Ser Tyr Ile Lys Ile Met Asp Val Gly Gln Ser Tyr
                215                 220                 225
Val Val Asn Arg Val Ala Asp His Ile Gln Ser Arg Ile Val Tyr
                230                 235                 240
Tyr Leu Met Asn Ile His Val Thr Pro Arg Ser Ile Tyr Leu Cys
```

PF-1506 P

```
                245                 250                 255
Arg His Gly Glu Ser Glu Leu Asn Leu Lys Gly Arg Ile Gly Gly
                260                 265                 270
Asp Pro Gly Leu Ser Pro Arg Gly Arg Glu Phe Ala Lys Ser Leu
                275                 280                 285
Ala Gln Phe Ile Ser Asp Gln Asn Ile Lys Asp Leu Lys Val Trp
                290                 295                 300
Thr Ser Gln Met Lys Arg Thr Ile Gln Thr Ala Glu Ala Leu Gly
                305                 310                 315
Val Pro Tyr Glu Gln Trp Lys Val Leu Asn Glu Ile Asp Ala Ser
                320                 325                 330
Tyr Glu Asp Leu Val Gln Arg Leu Glu Pro Val Ile Met Glu Leu
                335                 340                 345
Glu Arg Gln Glu Asn Val Leu Val Ile Cys His Gln Ala Val Met
                350                 355                 360
Arg Cys Leu Leu Ala Tyr Phe Leu Asp Lys Ala Ala Glu Gln Leu
                365                 370                 375
Pro Tyr Leu Lys Cys Pro Leu His Thr Val Leu Lys Leu Thr Pro
                380                 385                 390
Val Ala Tyr Gly Cys Lys Val Glu Ser Ile Phe Leu Asn Val Ala
                395                 400                 405
Ala Val Asn Thr His Arg Asp Arg Pro Gln Asn Val Asp Ile Ser
                410                 415                 420
Arg Pro Pro Glu Glu Ala Leu Val Thr Val Pro Ala His Gln
                425                 430
```

<210> 4
<211> 240
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7523965CD1

<400> 4

```
Met Ala Ala Leu Tyr Arg Pro Gly Leu Arg Leu Asn Trp His Gly
  1                 5                  10                  15
Leu Ser Pro Leu Gly Trp Pro Ser Cys Arg Ser Ile Gln Thr Leu
                 20                  25                  30
Arg Val Leu Ser Gly Asp Leu Gly Gln Leu Pro Thr Gly Ile Arg
                 35                  40                  45
Asp Phe Val Glu His Ser Ala Arg Leu Cys Gln Pro Glu Gly Ile
                 50                  55                  60
His Ile Cys Asp Gly Thr Glu Ala Glu Asn Thr Ala Thr Leu Thr
                 65                  70                  75
Leu Leu Glu Gln Gln Gly Leu Ile Arg Lys Leu Pro Lys Tyr Asn
                 80                  85                  90
Asn Cys Trp Leu Ala Arg Thr Asp Pro Lys Asp Val Ala Arg Val
                 95                 100                 105
Glu Ser Lys Thr Val Ile Val Thr Pro Ser Gln Arg Asp Thr Val
                110                 115                 120
Pro Leu Pro Pro Gly Gly Ala Arg Gly Gln Leu Gly Asn Trp Met
                125                 130                 135
Ser Pro Ala Asp Phe Gln Arg Ala Val Asp Glu Arg Phe Pro Gly
                140                 145                 150
Cys Met Gln Gly Arg Thr Met Tyr Val Leu Pro Phe Ser Met Gly
                155                 160                 165
Pro Val Gly Ser Pro Leu Ser Arg Ile Gly Val Gln Leu Thr Asp
                170                 175                 180
Ser Ala Tyr Val Val Ala Ser Met Arg Ile Met Thr Arg Leu Gly
                185                 190                 195
Thr Pro Val Leu Gln Ala Leu Gly Asp Gly Asp Phe Val Lys Cys
                200                 205                 210
Leu His Ser Val Gly Gln Pro Leu Thr Gly Gln Asp Pro Gly His
                215                 220                 225
```

PF-1506 P

```
His Gln Pro Cys Arg Glu Glu Ala Leu Cys Gly Ser Arg Leu Pro
                230                 235                 240
```

<210> 5
<211> 199
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7524016CD1

<400> 5

```
Met Glu Glu Lys Thr Ser Arg Ile Lys Ala Ser Ile Pro Gln Phe
  1               5                  10                  15
Thr Asn Ser Pro Thr Met Val Ile Met Val Gly Leu Pro Ala Arg
             20                  25                  30
Gly Lys Thr Tyr Ile Ser Thr Lys Leu Thr Arg Tyr Leu Asn Trp
             35                  40                  45
Ile Gly Thr Pro Thr Lys Val Phe Asn Leu Gly Gln Tyr Arg Arg
             50                  55                  60
Glu Ala Val Ser Tyr Lys Asn Tyr Glu Phe Phe Leu Pro Asp Asn
             65                  70                  75
Met Glu Ala Leu Gln Ile Arg Lys Gln Cys Ala Leu Ala Ala Leu
             80                  85                  90
Lys Asp Val His Asn Tyr Leu Ser His Glu Glu Gly His Val Ala
             95                 100                 105
Val Phe Asp Ala Thr Asn Thr Thr Arg Glu Arg Arg Ser Leu Ile
            110                 115                 120
Leu Gln Phe Ala Lys Glu His Gly Tyr Lys Val Phe Phe Ile Glu
            125                 130                 135
Ser Ile Cys Asn Asp Pro Gly Ile Ile Ala Glu Asn Ile Arg Gln
            140                 145                 150
Val Lys Leu Gly Ser Pro Asp Tyr Ile Asp Cys Asp Arg Glu Lys
            155                 160                 165
Val Leu Glu Asp Phe Leu Lys Arg Ile Glu Cys Tyr Glu Val Asn
            170                 175                 180
Tyr Gln Pro Leu Asp Glu Glu Leu Asp Arg Ser Ser Thr Trp Ala
            185                 190                 195
His Ala Thr Trp
```

<210> 6
<211> 406
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7524680CD1

<400> 6

```
Met Glu Glu Lys Thr Ser Arg Ile Lys Val Phe Asn Leu Gly Gln
  1               5                  10                  15
Tyr Arg Arg Glu Ala Val Ser Tyr Lys Asn Tyr Glu Phe Phe Leu
             20                  25                  30
Pro Asp Asn Met Glu Ala Leu Gln Ile Arg Lys Gln Cys Ala Leu
             35                  40                  45
Ala Ala Leu Lys Asp Val His Asn Tyr Leu Ser His Glu Glu Gly
             50                  55                  60
His Val Ala Val Phe Asp Ala Thr Asn Thr Thr Arg Glu Arg Arg
             65                  70                  75
Ser Leu Ile Leu Gln Phe Ala Lys Glu His Gly Tyr Lys Val Phe
             80                  85                  90
Phe Ile Glu Ser Ile Cys Asn Asp Pro Gly Ile Ile Ala Glu Asn
```

PF-1506 P

95 100 105
Ile Arg Gln Val Lys Leu Gly Ser Pro Asp Tyr Ile Asp Cys Asp
110 115 120
Arg Glu Lys Val Leu Glu Asp Phe Leu Lys Arg Ile Glu Cys Tyr
125 130 135
Glu Val Asn Tyr Gln Pro Leu Asp Glu Glu Leu Asp Ser His Leu
140 145 150
Ser Tyr Ile Lys Ile Phe Asp Val Gly Thr Arg Tyr Met Val Asn
155 160 165
Arg Val Gln Asp His Ile Gln Ser Arg Thr Val Tyr Tyr Leu Met
170 175 180
Asn Ile His Val Thr Pro Arg Ser Ile Tyr Leu Cys Arg His Gly
185 190 195
Glu Ser Glu Leu Asn Ile Arg Gly Arg Ile Gly Gly Asp Ser Gly
200 205 210
Leu Ser Val Arg Gly Lys Gln Tyr Ala Tyr Ala Leu Ala Asn Phe
215 220 225
Ile Gln Ser Gln Gly Ile Ser Ser Leu Lys Val Trp Thr Ser His
230 235 240
Met Lys Arg Thr Ile Gln Thr Ala Glu Ala Leu Gly Val Pro Tyr
245 250 255
Glu Gln Trp Lys Ala Leu Asn Glu Ile Asp Ala Gly Val Cys Glu
260 265 270
Glu Met Thr Tyr Glu Glu Ile Gln Glu His Tyr Pro Glu Glu Phe
275 280 285
Ala Leu Arg Asp Gln Asp Lys Tyr Arg Tyr Arg Tyr Pro Lys Gly
290 295 300
Glu Ser Tyr Glu Asp Leu Val Gln Arg Leu Glu Pro Val Ile Met
305 310 315
Glu Leu Glu Arg Gln Glu Asn Val Leu Val Ile Cys His Gln Ala
320 325 330
Val Met Arg Cys Leu Leu Ala Tyr Phe Leu Asp Lys Ser Ser Asp
335 340 345
Glu Leu Pro Tyr Leu Lys Cys Pro Leu His Thr Val Leu Lys Leu
350 355 360
Thr Pro Val Ala Tyr Gly Cys Lys Val Glu Ser Ile Tyr Leu Asn
365 370 375
Val Glu Thr Val Asn Thr His Arg Glu Lys Pro Glu Asn Val Asp
380 385 390
Ile Thr Arg Glu Pro Glu Glu Ala Leu Asp Thr Val Pro Ala His
395 400 405
Tyr

<210> 7
<211> 426
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7524757CD1

<400> 7
Met Glu Glu Lys Thr Ser Arg Ile Lys Ala Ser Ile Pro Gln Phe
1 5 10 15
Thr Asn Ser Pro Thr Met Val Ile Met Val Gly Leu Pro Ala Arg
20 25 30
Gly Lys Thr Tyr Ile Ser Thr Lys Leu Thr Arg Tyr Leu Asn Trp
35 40 45
Ile Gly Thr Pro Thr Lys Asp Asn Met Glu Ala Leu Gln Ile Arg
50 55 60
Lys Gln Cys Ala Leu Ala Ala Leu Lys Asp Val His Asn Tyr Leu
65 70 75
Ser His Glu Glu Gly His Val Ala Val Phe Asp Ala Thr Asn Thr
80 85 90

PF-1506 P

```
Thr Arg Glu Arg Arg Ser Leu Ile Leu Gln Phe Ala Lys Glu His
                 95                  100                 105
Gly Tyr Lys Val Phe Phe Ile Glu Ser Ile Cys Asn Asp Pro Gly
                110                 115                 120
Ile Ile Ala Glu Asn Ile Arg Gln Val Lys Leu Gly Ser Pro Asp
                125                 130                 135
Tyr Ile Asp Cys Asp Arg Glu Lys Val Leu Glu Asp Phe Leu Lys
                140                 145                 150
Arg Ile Glu Cys Tyr Glu Val Asn Tyr Gln Pro Leu Asp Glu Glu
                155                 160                 165
Leu Asp Ser His Leu Ser Tyr Ile Lys Ile Phe Asp Val Gly Thr
                170                 175                 180
Arg Tyr Met Val Asn Arg Val Gln Asp His Ile Gln Ser Arg Thr
                185                 190                 195
Val Tyr Tyr Leu Met Asn Ile His Val Thr Pro Arg Ser Ile Tyr
                200                 205                 210
Leu Cys Arg His Gly Glu Ser Glu Leu Asn Ile Arg Gly Arg Ile
                215                 220                 225
Gly Gly Asp Ser Gly Leu Ser Val Arg Gly Lys Gln Tyr Ala Tyr
                230                 235                 240
Ala Leu Ala Asn Phe Ile Gln Ser Gln Gly Ile Ser Ser Leu Lys
                245                 250                 255
Val Trp Thr Ser His Met Lys Arg Thr Ile Gln Thr Ala Glu Ala
                260                 265                 270
Leu Gly Val Pro Tyr Glu Gln Trp Lys Ala Leu Asn Glu Ile Asp
                275                 280                 285
Ala Gly Val Cys Glu Glu Met Thr Tyr Glu Glu Ile Arg Glu His
                290                 295                 300
Tyr Pro Glu Glu Phe Ala Leu Arg Asp Gln Asp Lys Tyr Arg Tyr
                305                 310                 315
Arg Tyr Pro Lys Gly Glu Ser Tyr Glu Asp Leu Val Gln Arg Leu
                320                 325                 330
Glu Pro Val Ile Met Glu Leu Glu Arg Gln Glu Asn Val Leu Val
                335                 340                 345
Ile Cys His Gln Ala Val Met Arg Cys Leu Leu Ala Tyr Phe Leu
                350                 355                 360
Asp Lys Ser Ser Asp Glu Leu Pro Tyr Leu Lys Cys Pro Leu His
                365                 370                 375
Thr Val Leu Lys Leu Thr Pro Val Ala Tyr Gly Cys Lys Val Glu
                380                 385                 390
Ser Ile Tyr Leu Asn Val Glu Ala Val Asn Thr His Arg Glu Lys
                395                 400                 405
Pro Glu Asn Val Asp Ile Thr Arg Glu Pro Glu Glu Ala Leu Asp
                410                 415                 420
Thr Val Pro Ala His Tyr
                425
```

<210> 8
<211> 355
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7516229CD1

<400> 8

```
Met Ala Thr Pro Gly Asn Leu Gly Ser Ser Val Leu Ala Ser Lys
  1               5                  10                  15
Thr Lys Thr Lys Lys Lys His Phe Val Ala Gln Lys Val Lys Leu
                 20                  25                  30
Phe Arg Ala Ser Asp Pro Leu Leu Ser Val Leu Met Trp Gly Val
                 35                  40                  45
Asn His Ser Ile Asn Glu Leu Ser His Val Gln Ile Pro Val Met
                 50                  55                  60
Leu Met Pro Asp Asp Phe Lys Ala Tyr Ser Lys Ile Lys Val Asp
```

PF-1506 P

```
                  65                  70                  75
Asn His Leu Phe Asn Lys Glu Asn Met Pro Ser His Phe Lys Phe
                  80                  85                  90
Lys Glu Tyr Cys Pro Met Val Phe Arg Asn Leu Arg Glu Arg Phe
                  95                 100                 105
Gly Ile Asp Asp Gln Asp Phe Gln Tyr Ile Val Glu Cys His Gly
                 110                 115                 120
Ile Thr Leu Leu Pro Gln Phe Leu Gly Met Tyr Arg Leu Asn Val
                 125                 130                 135
Asp Gly Val Glu Ile Tyr Val Ile Val Thr Arg Asn Val Phe Ser
                 140                 145                 150
His Arg Leu Ser Val Tyr Arg Lys Tyr Asp Leu Lys Gly Ser Thr
                 155                 160                 165
Val Ala Arg Glu Ala Ser Asp Lys Glu Lys Ala Lys Glu Leu Pro
                 170                 175                 180
Thr Leu Lys Asp Asn Asp Phe Ile Asn Glu Gly Gln Lys Ile Tyr
                 185                 190                 195
Ile Asp Asp Asn Asn Lys Lys Val Phe Leu Glu Lys Leu Lys Lys
                 200                 205                 210
Asp Val Glu Phe Leu Ala Gln Leu Lys Leu Met Asp Tyr Ser Leu
                 215                 220                 225
Leu Val Gly Ile His Asp Val Glu Arg Ala Glu Gln Glu Glu Val
                 230                 235                 240
Glu Cys Glu Glu Asn Asp Gly Glu Glu Glu Gly Glu Ser Asp Gly
                 245                 250                 255
Thr His Pro Val Gly Thr Pro Pro Asp Ser Pro Gly Asn Thr Leu
                 260                 265                 270
Asn Ser Ser Pro Pro Leu Ala Pro Gly Glu Phe Asp Pro Asn Ile
                 275                 280                 285
Asp Val Tyr Gly Ile Lys Cys His Glu Asn Ser Pro Arg Lys Glu
                 290                 295                 300
Val Tyr Phe Met Ala Ile Ile Asp Ile Leu Thr His Tyr Asp Ala
                 305                 310                 315
Lys Lys Lys Ala Ala His Ala Ala Lys Thr Val Lys His Gly Ala
                 320                 325                 330
Gly Ala Glu Ile Ser Thr Val Asn Pro Glu Gln Tyr Ser Lys Arg
                 335                 340                 345
Phe Leu Asp Phe Ile Gly His Ile Leu Thr
                 350                 355
```

<210> 9
<211> 543
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7516525CD1

<400> 9

```
Met Glu Gly Gly Pro Ala Val Cys Cys Gln Asp Pro Arg Ala Glu
  1               5                  10                  15
Leu Val Glu Arg Val Ala Ala Ile Asp Val Thr His Leu Glu Glu
                 20                  25                  30
Ala Asp Gly Gly Pro Glu Pro Thr Arg Asn Gly Val Asp Pro Pro
                 35                  40                  45
Pro Arg Ala Arg Ala Ala Ser Val Ile Pro Gly Ser Thr Ser Arg
                 50                  55                  60
Leu Leu Pro Ala Arg Pro Ser Leu Ser Ala Arg Lys Leu Ser Leu
                 65                  70                  75
Gln Glu Arg Pro Ala Gly Ser Tyr Leu Glu Ala Gln Ala Gly Pro
                 80                  85                  90
Tyr Ala Thr Gly Pro Ala Ser His Ile Ser Pro Arg Ala Trp Arg
                 95                 100                 105
Arg Pro Thr Ile Glu Ser His His Val Ala Ile Ser Asp Ala Glu
                110                 115                 120
```

PF-1506 P

```
Asp Cys Val Gln Leu Asn Gln Tyr Lys Leu Gln Ser Glu Ile Gly
                125                 130                 135
Lys Gly Ala Tyr Gly Val Val Arg Pro Ala Tyr Asn Glu Ser Glu
            140                 145                 150
Asp Arg His Tyr Ala Met Lys Val Leu Ser Lys Lys Lys Leu Leu
        155                 160                 165
Lys Gln Tyr Gly Phe Pro Arg Arg Pro Pro Pro Arg Gly Ser Gln
    170                 175                 180
Ala Ala Gln Gly Gly Pro Ala Lys Gln Leu Leu Pro Leu Glu Arg
185                 190                 195
Val Tyr Gln Glu Ile Ala Ile Leu Lys Lys Leu Asp His Val Asn
            200                 205                 210
Val Val Lys Leu Ile Glu Val Leu Asp Asp Pro Ala Glu Asp Asn
        215                 220                 225
Leu Tyr Leu Ala Leu Gln Asn Gln Ala Gln Asn Ile Gln Leu Asp
    230                 235                 240
Ser Thr Asn Ile Ala Lys Pro His Ser Leu Leu Pro Ser Glu Gln
245                 250                 255
Gln Asp Ser Gly Ser Thr Trp Ala Ala Arg Ser Val Phe Asp Leu
            260                 265                 270
Leu Arg Lys Gly Pro Val Met Glu Val Pro Cys Asp Lys Pro Phe
        275                 280                 285
Ser Glu Glu Gln Ala Arg Leu Tyr Leu Arg Asp Val Ile Leu Gly
    290                 295                 300
Leu Glu Tyr Leu His Cys Gln Lys Ile Val His Arg Asp Ile Lys
305                 310                 315
Pro Ser Asn Leu Leu Leu Gly Asp Asp Gly His Val Lys Ile Ala
            320                 325                 330
Asp Phe Gly Val Ser Asn Gln Phe Glu Gly Asn Asp Ala Gln Leu
        335                 340                 345
Ser Ser Thr Ala Gly Thr Pro Ala Phe Met Ala Pro Glu Ala Ile
    350                 355                 360
Ser Asp Ser Gly Gln Ser Phe Ser Gly Lys Ala Leu Asp Val Trp
365                 370                 375
Ala Thr Gly Val Thr Leu Tyr Cys Phe Val Tyr Gly Lys Cys Pro
            380                 385                 390
Phe Ile Asp Asp Phe Ile Leu Ala Leu His Arg Lys Ile Lys Asn
        395                 400                 405
Glu Pro Val Val Phe Pro Glu Gly Pro Glu Ile Ser Glu Glu Leu
    410                 415                 420
Lys Asp Leu Ile Leu Lys Met Leu Asp Lys Asn Pro Glu Thr Arg
425                 430                 435
Ile Gly Val Pro Asp Ile Lys Leu His Pro Trp Val Thr Lys Asn
            440                 445                 450
Gly Glu Glu Pro Ile Pro Ser Glu Glu Glu His Cys Ser Val Val
        455                 460                 465
Glu Val Thr Glu Glu Glu Val Lys Asn Ser Val Arg Leu Ile Pro
    470                 475                 480
Ser Trp Thr Thr Val Ile Leu Val Lys Ser Met Leu Arg Lys Arg
485                 490                 495
Ser Phe Gly Asn Pro Phe Glu Pro Gln Ala Arg Arg Glu Glu Arg
            500                 505                 510
Ser Met Ser Ala Pro Gly Asn Leu Leu Val Lys Glu Gly Phe Gly
        515                 520                 525
Glu Gly Gly Lys Ser Pro Glu Leu Pro Gly Val Gln Glu Asp Glu
    530                 535                 540
Ala Ala Ser
```

<210> 10
<211> 445
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

PF-1506 P

<223> Incyte ID No: 7516533CD1

<400> 10

```
Met Arg Arg Arg Arg Arg Arg Asp Gly Phe Tyr Pro Ala Pro Asp
  1               5                  10                  15
Phe Arg Asp Arg Glu Ala Glu Asp Met Ala Gly Val Phe Asp Ile
             20                  25                  30
Asp Leu Asp Gln Pro Glu Asp Ala Gly Ser Glu Asp Glu Leu Glu
         35                  40                  45
Glu Gly Ala Met Ile Val Arg Asn Ala Lys Asp Thr Ala His Thr
     50                  55                  60
Lys Ala Glu Arg Asn Ile Leu Glu Glu Val Lys His Pro Phe Ile
 65                  70                  75
Val Asp Leu Ile Tyr Ala Phe Gln Thr Gly Gly Lys Leu Tyr Leu
             80                  85                  90
Ile Leu Glu Tyr Leu Ser Gly Gly Glu Leu Phe Met Gln Leu Glu
         95                 100                 105
Arg Glu Gly Ile Phe Met Glu Asp Thr Ala Cys Phe Tyr Leu Ala
    110                 115                 120
Glu Ile Ser Met Ala Leu Gly His Leu His Gln Lys Gly Ile Ile
125                 130                 135
Tyr Arg Asp Leu Lys Pro Glu Asn Ile Met Leu Asn His Gln Gly
140                 145                 150
His Val Lys Leu Thr Asp Phe Gly Leu Cys Lys Glu Ser Ile His
155                 160                 165
Asp Gly Thr Val Thr His Thr Phe Cys Gly Thr Ile Glu Tyr Met
170                 175                 180
Ala Pro Glu Ile Leu Met Arg Ser Gly His Asn Arg Ala Val Asp
185                 190                 195
Trp Trp Ser Leu Gly Ala Leu Met Tyr Asp Met Leu Thr Gly Ala
200                 205                 210
Pro Pro Phe Thr Gly Glu Asn Arg Lys Lys Thr Ile Asp Lys Ile
215                 220                 225
Leu Lys Cys Lys Leu Asn Leu Pro Pro Tyr Leu Thr Gln Glu Ala
230                 235                 240
Arg Asp Leu Leu Lys Lys Leu Leu Lys Arg Asn Ala Ala Ser Arg
245                 250                 255
Leu Gly Ala Gly Pro Gly Asp Ala Gly Glu Val Gln Ala His Pro
260                 265                 270
Phe Phe Arg His Ile Asn Trp Glu Glu Leu Leu Ala Arg Lys Val
275                 280                 285
Glu Pro Pro Phe Lys Pro Leu Leu Gln Ser Glu Glu Asp Val Ser
290                 295                 300
Gln Phe Asp Ser Lys Phe Thr Arg Gln Thr Pro Val Asp Ser Pro
305                 310                 315
Asp Asp Ser Thr Leu Ser Glu Ser Ala Asn Gln Val Phe Leu Gly
320                 325                 330
Phe Thr Tyr Val Ala Pro Ser Val Leu Glu Ser Val Lys Glu Lys
335                 340                 345
Phe Ser Phe Glu Pro Lys Ile Arg Ser Pro Arg Arg Phe Ile Gly
350                 355                 360
Ser Pro Arg Thr Pro Val Ser Pro Val Lys Phe Ser Pro Gly Asp
365                 370                 375
Phe Trp Gly Arg Gly Ala Ser Ala Ser Ala Ala Asn Pro Gln Thr
380                 385                 390
Pro Val Glu Tyr Pro Met Glu Thr Ser Gly Ile Glu Gln Met Asp
395                 400                 405
Val Thr Met Ser Gly Glu Ala Ser Ala Pro Leu Pro Ile Arg Gln
410                 415                 420
Pro Asn Ser Gly Pro Tyr Lys Lys Gln Ala Phe Pro Met Ile Ser
425                 430                 435
Lys Arg Pro Glu His Leu Arg Met Asn Leu
440                 445
```

<210> 11
<211> 1219

PF-1506 P

<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7516613CD1

<400> 11

```
Met Ala Asn Asp Ser Pro Ala Lys Ser Leu Val Asp Ile Asp Leu
  1               5                  10                  15
Ser Ser Leu Arg Asp Pro Ala Gly Ile Phe Glu Leu Val Glu Val
             20                  25                  30
Val Gly Asn Gly Thr Tyr Gly Gln Val Tyr Lys Gly Arg His Val
         35                  40                  45
Lys Thr Gly Gln Leu Ala Ala Ile Lys Val Met Asp Val Thr Glu
     50                  55                  60
Asp Glu Glu Glu Glu Ile Lys Leu Glu Ile Asn Met Leu Lys Lys
 65                  70                  75
Tyr Ser His His Arg Asn Ile Ala Thr Tyr Tyr Gly Ala Phe Ile
             80                  85                  90
Lys Lys Ser Pro Pro Gly His Asp Asp Gln Leu Trp Leu Val Met
         95                 100                 105
Glu Phe Cys Gly Ala Gly Ser Ile Thr Asp Leu Val Lys Asn Thr
    110                 115                 120
Lys Gly Asn Thr Leu Lys Glu Asp Trp Ile Ala Tyr Ile Ser Arg
125                 130                 135
Glu Ile Leu Arg Gly Leu Ala His Leu His Ile His His Val Ile
            140                 145                 150
His Arg Asp Ile Lys Gly Gln Asn Val Leu Leu Thr Glu Asn Ala
        155                 160                 165
Glu Val Lys Leu Val Asp Phe Gly Val Ser Ala Gln Leu Asp Gly
    170                 175                 180
Thr Val Gly Arg Arg Asn Thr Phe Ile Gly Thr Pro Tyr Trp Met
185                 190                 195
Ala Pro Glu Val Ile Ala Cys Asp Glu Asn Pro Asp Ala Thr Tyr
            200                 205                 210
Asp Tyr Arg Ser Asp Leu Trp Ser Cys Gly Ile Thr Ala Ile Glu
        215                 220                 225
Met Gly Glu Gly Ala Pro Pro Leu Cys Asp Met His Pro Met Arg
    230                 235                 240
Ala Leu Phe Leu Ile Pro Arg Asn Pro Pro Pro Arg Leu Lys Ser
245                 250                 255
Lys Lys Trp Ser Lys Lys Phe Phe Ser Phe Ile Glu Gly Cys Leu
            260                 265                 270
Val Lys Asn Tyr Met Gln Arg Pro Ser Thr Glu Gln Leu Leu Lys
        275                 280                 285
His Pro Phe Ile Arg Asp Gln Pro Asn Glu Arg Gln Val Arg Ile
    290                 295                 300
Gln Leu Lys Asp His Ile Asp Arg Thr Arg Lys Lys Arg Gly Glu
305                 310                 315
Lys Asp Glu Thr Glu Tyr Glu Tyr Ser Gly Ser Glu Glu Glu Glu
            320                 325                 330
Glu Glu Val Pro Glu Gln Glu Gly Glu Pro Ser Ser Ile Val Asn
        335                 340                 345
Val Pro Gly Glu Ser Thr Leu Arg Arg Asp Phe Leu Arg Leu Gln
    350                 355                 360
Gln Glu Asn Lys Glu Arg Ser Glu Ala Leu Arg Arg Gln Gln Leu
365                 370                 375
Leu Gln Glu Gln Gln Leu Arg Glu Gln Glu Glu Tyr Lys Arg Gln
            380                 385                 390
Leu Leu Ala Glu Arg Gln Lys Arg Ile Glu Gln Gln Lys Glu Gln
        395                 400                 405
Arg Arg Arg Leu Glu Glu Gln Gln Arg Arg Glu Arg Glu Ala Arg
    410                 415                 420
Arg Gln Gln Glu Arg Glu Gln Arg Arg Arg Glu Gln Glu Glu Lys
425                 430                 435
```

PF-1506 P

```
Arg Arg Leu Glu Glu Leu Glu Arg Arg Arg Lys Glu Glu Glu Glu
                440                 445                 450
Arg Arg Gln Ala Glu Glu Glu Lys Arg Arg Val Glu Arg Glu Gln
                455                 460                 465
Glu Tyr Ile Arg Arg Gln Leu Glu Glu Glu Gln Arg His Leu Glu
                470                 475                 480
Val Leu Gln Gln Gln Leu Leu Gln Glu Gln Ala Met Leu Leu His
                485                 490                 495
Asp His Arg Arg Pro His Pro Gln His Ser Gln Gln Pro Pro Pro
                500                 505                 510
Pro Gln Gln Glu Arg Ser Lys Pro Ser Phe His Ala Pro Glu Pro
                515                 520                 525
Lys Ala His Tyr Glu Pro Ala Asp Arg Ala Arg Glu Val Pro Val
                530                 535                 540
Arg Thr Thr Ser Arg Ser Pro Val Leu Ser Arg Arg Asp Ser Pro
                545                 550                 555
Leu Gln Gly Ser Gly Gln Gln Asn Ser Gln Ala Gly Gln Arg Asn
                560                 565                 570
Ser Thr Ser Ser Ile Glu Pro Arg Leu Leu Trp Glu Arg Val Glu
                575                 580                 585
Lys Leu Met Pro Arg Pro Gly Ser Gly Ser Ser Ser Gly Ser Ser
                590                 595                 600
Asn Ser Gly Ser Gln Pro Gly Ser His Pro Gly Ser Gln Ser Gly
                605                 610                 615
Ser Gly Glu Arg Phe Arg Val Arg Ser Ser Ser Lys Ser Glu Gly
                620                 625                 630
Ser Pro Ser Gln Arg Leu Glu Asn Ala Val Lys Lys Pro Glu Asp
                635                 640                 645
Lys Lys Glu Val Phe Arg Pro Leu Lys Pro Ala Asp Leu Thr Ala
                650                 655                 660
Leu Ala Lys Glu Leu Arg Ala Val Glu Asp Val Arg Pro Pro His
                665                 670                 675
Lys Val Thr Asp Tyr Ser Ser Ser Ser Glu Glu Pro Gly Thr Thr
                680                 685                 690
Asp Glu Glu Asp Asp Asp Val Glu Gln Glu Gly Ala Asp Glu Ser
                695                 700                 705
Thr Ser Gly Pro Glu Asp Thr Arg Ala Ala Ser Ser Leu Asn Leu
                710                 715                 720
Ser Asn Gly Glu Thr Glu Ser Val Lys Thr Met Ile Val His Asp
                725                 730                 735
Asp Val Glu Ser Glu Pro Ala Met Thr Pro Ser Lys Glu Gly Thr
                740                 745                 750
Leu Ile Val Arg Gln Ser Thr Val Asp Gln Lys Arg Ala Ser His
                755                 760                 765
His Glu Ser Asn Gly Phe Ala Gly Arg Ile His Leu Leu Pro Asp
                770                 775                 780
Leu Leu Gln Gln Ser His Ser Ser Ser Thr Ser Ser Thr Ser Ser
                785                 790                 795
Ser Pro Ser Ser Ser Gln Pro Thr Pro Thr Met Ser Pro Gln Thr
                800                 805                 810
Pro Gln Asp Lys Leu Thr Ala Asn Glu Thr Gln Ser Ala Ser Ser
                815                 820                 825
Thr Leu Gln Lys His Lys Ser Ser Ser Ser Phe Thr Pro Phe Ile
                830                 835                 840
Asp Pro Arg Leu Leu Gln Ile Ser Pro Ser Ser Gly Thr Thr Val
                845                 850                 855
Thr Ser Val Val Gly Phe Ser Cys Asp Gly Met Arg Pro Glu Ala
                860                 865                 870
Ile Arg Gln Asp Pro Thr Arg Lys Gly Ser Val Val Asn Val Asn
                875                 880                 885
Pro Thr Asn Thr Arg Pro Gln Ser Asp Thr Pro Glu Ile Arg Lys
                890                 895                 900
Tyr Lys Lys Arg Phe Asn Ser Glu Ile Leu Cys Ala Ala Leu Trp
                905                 910                 915
Gly Val Asn Leu Leu Val Gly Thr Glu Ser Gly Leu Met Leu Leu
                920                 925                 930
```

PF-1506 P

```
Asp Arg Ser Gly Gln Gly Lys Val Tyr Pro Leu Ile Asn Arg Arg
                935                 940                 945
Arg Phe Gln Gln Met Asp Val Leu Glu Gly Leu Asn Val Leu Val
                950                 955                 960
Thr Ile Ser Gly Lys Lys Asp Lys Leu Arg Val Tyr Tyr Leu Ser
                965                 970                 975
Trp Leu Arg Asn Lys Ile Leu His Asn Asp Pro Glu Val Glu Lys
                980                 985                 990
Lys Gln Gly Trp Thr Thr Val Gly Asp Leu Glu Gly Cys Val His
                995                1000                1005
Tyr Lys Val Val Lys Tyr Glu Arg Ile Lys Phe Leu Val Ile Ala
               1010                1015                1020
Leu Lys Ser Ser Val Glu Val Tyr Ala Trp Ala Pro Lys Pro Tyr
               1025                1030                1035
His Lys Phe Met Ala Phe Lys Ser Phe Gly Glu Leu Val His Lys
               1040                1045                1050
Pro Leu Leu Val Asp Leu Thr Val Glu Glu Gly Gln Arg Leu Lys
               1055                1060                1065
Val Ile Tyr Gly Ser Cys Ala Gly Phe His Ala Val Asp Val Asp
               1070                1075                1080
Ser Gly Ser Val Tyr Asp Ile Tyr Leu Pro Thr His Ile Gln Cys
               1085                1090                1095
Ser Ile Lys Pro His Ala Ile Ile Ile Leu Pro Asn Thr Asp Gly
               1100                1105                1110
Met Glu Leu Leu Val Cys Tyr Glu Asp Glu Gly Val Tyr Val Asn
               1115                1120                1125
Thr Tyr Gly Arg Ile Thr Lys Asp Val Val Leu Gln Trp Gly Glu
               1130                1135                1140
Met Pro Thr Ser Val Ala Tyr Ile Arg Ser Asn Gln Thr Met Gly
               1145                1150                1155
Trp Gly Glu Lys Ala Ile Glu Ile Arg Ser Val Glu Thr Gly His
               1160                1165                1170
Leu Asp Gly Val Phe Met His Lys Arg Ala Gln Arg Leu Lys Phe
               1175                1180                1185
Leu Cys Glu Arg Asn Asp Lys Val Phe Phe Ala Ser Val Arg Ser
               1190                1195                1200
Gly Gly Ser Ser Gln Val Tyr Phe Met Thr Leu Gly Arg Thr Ser
               1205                1210                1215
Leu Leu Ser Trp
```

<210> 12
<211> 1168
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7517068CD1

<400> 12

```
Met Ala Ser Asp Ser Pro Ala Arg Ser Leu Asp Glu Ile Asp Leu
  1                 5                  10                  15
Ser Ala Leu Arg Asp Pro Ala Gly Ile Phe Glu Leu Val Glu Leu
                20                  25                  30
Val Gly Asn Gly Thr Tyr Gly Gln Val Tyr Lys Gly Arg His Val
                35                  40                  45
Lys Thr Gly Gln Leu Ala Ala Ile Lys Val Met Asp Val Thr Gly
                50                  55                  60
Asp Glu Glu Glu Glu Ile Lys Gln Glu Ile Asn Met Leu Lys Lys
                65                  70                  75
Tyr Ser His His Arg Asn Ile Ala Thr Tyr Tyr Gly Ala Phe Ile
                80                  85                  90
Lys Lys Asn Pro Pro Gly Met Asp Asp Gln Leu Trp Leu Val Met
                95                 100                 105
Glu Phe Cys Gly Ala Gly Ser Val Thr Asp Leu Ile Lys Asn Thr
```

```
                110                 115                 120
Lys Gly Asn Thr Leu Lys Glu Glu Trp Ile Ala Tyr Ile Cys Arg
            125                 130                 135
Glu Ile Leu Arg Gly Leu Ser His Leu His Gln His Lys Val Ile
        140                 145                 150
His Arg Asp Ile Lys Gly Gln Asn Val Leu Leu Thr Glu Asn Ala
    155                 160                 165
Glu Val Lys Leu Val Asp Phe Gly Val Ser Ala Gln Leu Asp Arg
170                 175                 180
Thr Val Gly Arg Arg Asn Thr Phe Ile Gly Thr Pro Tyr Trp Met
185                 190                 195
Ala Pro Glu Val Ile Ala Cys Asp Glu Asn Pro Asp Ala Thr Tyr
200                 205                 210
Asp Phe Lys Ser Asp Leu Trp Ser Leu Gly Ile Thr Ala Ile Glu
215                 220                 225
Met Ala Glu Gly Ala Pro Pro Leu Cys Asp Met His Pro Met Arg
230                 235                 240
Ala Leu Phe Leu Ile Pro Arg Asn Pro Ala Pro Arg Leu Lys Ser
245                 250                 255
Lys Lys Trp Ser Lys Lys Phe Gln Ser Phe Ile Glu Ser Cys Leu
260                 265                 270
Val Lys Asn His Ser Gln Arg Pro Ala Thr Glu Gln Leu Met Lys
275                 280                 285
His Pro Phe Ile Arg Asp Gln Pro Asn Glu Arg Gln Val Arg Ile
290                 295                 300
Gln Leu Lys Asp His Ile Asp Arg Thr Lys Lys Lys Arg Gly Glu
305                 310                 315
Lys Asp Glu Thr Glu Tyr Glu Tyr Ser Gly Ser Glu Glu Glu Glu
320                 325                 330
Glu Glu Asn Asp Ser Gly Glu Pro Ser Ser Ile Leu Asn Leu Pro
335                 340                 345
Gly Glu Ser Thr Leu Arg Arg Asp Phe Leu Arg Leu Gln Leu Ala
350                 355                 360
Asn Lys Glu Arg Ser Glu Ala Leu Arg Arg Gln Gln Leu Glu Gln
365                 370                 375
Gln Gln Arg Glu Asn Glu Glu His Lys Arg Gln Leu Leu Ala Glu
380                 385                 390
Arg Gln Lys Arg Ile Glu Glu Gln Lys Glu Gln Arg Arg Arg Leu
395                 400                 405
Glu Glu Ile Pro His Leu Val Ala Val Lys Ser Gln Gly Pro Ala
410                 415                 420
Leu Thr Ala Ser Gln Ser Val His Glu Gln Pro Thr Lys Gly Leu
425                 430                 435
Ser Gly Phe Gln Glu Ala Leu Asn Val Thr Ser His Arg Val Glu
440                 445                 450
Met Pro Arg Gln Asn Ser Asp Pro Thr Ser Glu Asn Pro Pro Leu
455                 460                 465
Pro Thr Arg Ile Glu Lys Phe Asp Arg Ser Ser Trp Leu Arg Gln
470                 475                 480
Glu Glu Asp Ile Pro Pro Lys Val Pro Gln Arg Thr Thr Ser Ile
485                 490                 495
Ser Pro Ala Leu Ala Arg Lys Asn Ser Pro Gly Asn Gly Ser Ala
500                 505                 510
Leu Gly Pro Arg Leu Gly Ser Gln Pro Ile Arg Ala Ser Asn Pro
515                 520                 525
Asp Leu Arg Arg Thr Glu Pro Ile Leu Glu Ser Pro Leu Gln Arg
530                 535                 540
Thr Ser Ser Gly Ser Ser Ser Ser Ser Ser Thr Pro Ser Ser Gln
545                 550                 555
Pro Ser Ser Gln Gly Gly Ser Gln Pro Gly Ser Gln Ala Gly Ser
560                 565                 570
Ser Gly Arg Thr Arg Val Arg Ala Asn Ser Lys Ser Glu Gly Ser
575                 580                 585
Pro Val Leu Pro His Glu Pro Ala Lys Val Lys Pro Glu Glu Ser
590                 595                 600
Arg Asp Ile Thr Arg Pro Ser Arg Pro Ala Asp Leu Thr Ala Leu
```

PF-1506 P

```
                605                 610                 615
Ala Lys Glu Leu Arg Glu Leu Arg Ile Glu Glu Thr Asn Arg Pro
                620                 625                 630
Met Lys Lys Val Thr Asp Tyr Ser Ser Ser Ser Glu Glu Ser Glu
                635                 640                 645
Ser Ser Glu Glu Glu Glu Glu Asp Gly Glu Ser Glu Thr His Asp
                650                 655                 660
Gly Thr Val Ala Val Ser Asp Ile Pro Arg Leu Ile Pro Thr Gly
                665                 670                 675
Ala Pro Gly Ser Asn Glu Gln Tyr Asn Val Gly Met Val Gly Thr
                680                 685                 690
His Gly Leu Glu Thr Ser His Ala Asp Ser Phe Ser Gly Ser Ile
                695                 700                 705
Ser Arg Glu Gly Thr Leu Met Ile Arg Glu Thr Ser Gly Glu Lys
                710                 715                 720
Lys Arg Ser Gly His Ser Asp Ser Asn Gly Phe Ala Gly His Ile
                725                 730                 735
Asn Leu Pro Asp Leu Val Gln Gln Ser His Ser Pro Ala Gly Thr
                740                 745                 750
Pro Thr Glu Gly Leu Gly Arg Val Ser Thr His Ser Gln Glu Met
                755                 760                 765
Asp Ser Gly Thr Glu Tyr Gly Met Gly Ser Ser Thr Lys Ala Ser
                770                 775                 780
Phe Thr Pro Phe Val Asp Pro Arg Val Tyr Gln Thr Ser Pro Thr
                785                 790                 795
Asp Glu Asp Glu Glu Asp Glu Glu Ser Ser Ala Ala Ala Leu Phe
                800                 805                 810
Thr Ser Glu Leu Leu Arg Gln Glu Gln Ala Lys Leu Asn Glu Ala
                815                 820                 825
Arg Lys Ile Ser Val Val Asn Val Asn Pro Thr Asn Ile Arg Pro
                830                 835                 840
His Ser Asp Thr Pro Glu Ile Arg Gln Tyr Lys Lys Arg Phe Asn
                845                 850                 855
Ser Glu Ile Leu Cys Ala Ala Leu Trp Gly Val Asn Leu Leu Val
                860                 865                 870
Gly Thr Glu Asn Gly Leu Met Leu Leu Asp Arg Ser Gly Gln Gly
                875                 880                 885
Lys Val Tyr Asn Leu Ile Asn Arg Arg Arg Phe Gln Gln Met Asp
                890                 895                 900
Val Leu Glu Gly Leu Asn Val Leu Val Thr Ile Ser Gly Lys Lys
                905                 910                 915
Asn Lys Leu Arg Val Tyr Tyr Leu Ser Trp Leu Arg Asn Arg Ile
                920                 925                 930
Leu His Asn Asp Pro Glu Val Glu Lys Lys Gln Gly Trp Ile Thr
                935                 940                 945
Val Gly Asp Leu Glu Gly Cys Ile His Tyr Lys Val Val Lys Tyr
                950                 955                 960
Glu Arg Ile Lys Phe Leu Val Ile Ala Leu Lys Asn Ala Val Glu
                965                 970                 975
Ile Tyr Ala Trp Ala Pro Lys Pro Tyr His Lys Phe Met Ala Phe
                980                 985                 990
Lys Ser Phe Ala Asp Leu Gln His Lys Pro Leu Leu Val Asp Leu
                995                1000                1005
Thr Val Glu Glu Gly Gln Arg Leu Lys Val Ile Phe Gly Ser His
               1010                1015                1020
Thr Gly Phe His Val Ile Asp Val Asp Ser Gly Asn Ser Tyr Asp
               1025                1030                1035
Ile Tyr Ile Pro Ser His Ile Gln Gly Asn Ile Thr Pro His Ala
               1040                1045                1050
Ile Val Ile Leu Pro Lys Thr Asp Gly Met Glu Met Leu Val Cys
               1055                1060                1065
Tyr Glu Asp Glu Gly Val Tyr Val Asp Thr Tyr Gly Arg Ile Thr
               1070                1075                1080
Lys Asp Val Val Leu Gln Trp Gly Glu Met Pro Thr Ser Val Ala
               1085                1090                1095
Tyr Ile His Ser Asp Gln Ile Met Gly Trp Gly Glu Lys Ala Ile
```

PF-1506 P

```
                1100                1105                1110
Glu Ile Arg Ser Val Glu Thr Gly His Leu Asp Gly Val Phe Met
                1115                1120                1125
His Lys Arg Ala Gln Arg Leu Lys Phe Leu Cys Glu Arg Asn Asp
                1130                1135                1140
Lys Val Phe Phe Ala Ser Val Arg Ser Gly Gly Ser Ser Gln Val
                1145                1150                1155
Phe Phe Met Thr Leu Asn Arg Asn Ser Met Met Asn Trp
                1160                1165
```

<210> 13
<211> 650
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7517148CD1

<400> 13

```
Met Ala Asp Leu Glu Ala Val Leu Ala Asp Val Ser Tyr Leu Met
  1               5                  10                  15
Ala Met Glu Lys Ser Lys Ala Thr Pro Ala Ala Arg Ala Ser Lys
                20                  25                  30
Arg Ile Val Leu Pro Glu Pro Ser Ile Arg Ser Val Met Gln Lys
                35                  40                  45
Tyr Leu Ala Glu Arg Asn Glu Ile Thr Leu Asp Lys Ile Phe Asn
                50                  55                  60
Gln Lys Ile Gly Phe Leu Leu Phe Lys Asp Phe Cys Leu Asn Glu
                65                  70                  75
Ile Asn Glu Ala Val Pro Gln Val Lys Phe Tyr Glu Glu Ile Lys
                80                  85                  90
Glu Tyr Glu Lys Leu Asp Asn Glu Glu Asp Arg Leu Cys Arg Ser
                95                 100                 105
Arg Gln Ile Tyr Asp Ala Tyr Ile Met Lys Glu Leu Leu Ser Cys
               110                 115                 120
Ser His Pro Phe Ser Lys Gln Ala Val Glu His Val Gln Ser His
               125                 130                 135
Leu Ser Lys Lys Gln Val Thr Ser Thr Leu Phe Gln Pro Tyr Ile
               140                 145                 150
Glu Glu Ile Cys Glu Ser Leu Arg Gly Asp Ile Phe Gln Lys Phe
               155                 160                 165
Met Glu Ser Asp Lys Phe Thr Arg Phe Cys Gln Trp Lys Asn Val
               170                 175                 180
Glu Leu Asn Ile His Leu Thr Met Asn Glu Phe Ser Val His Arg
               185                 190                 195
Ile Ile Gly Arg Gly Gly Phe Gly Glu Val Tyr Gly Cys Arg Lys
               200                 205                 210
Ala Asp Thr Gly Lys Met Tyr Ala Met Lys Cys Leu Asp Lys Lys
               215                 220                 225
Arg Ile Lys Met Lys Gln Gly Glu Thr Leu Ala Leu Asn Glu Arg
               230                 235                 240
Ile Met Leu Ser Leu Val Ser Thr Gly Asp Cys Pro Phe Ile Val
               245                 250                 255
Cys Met Thr Tyr Ala Phe His Thr Pro Asp Lys Leu Cys Phe Ile
               260                 265                 270
Leu Asp Leu Met Asn Gly Gly Asp Leu His Tyr His Leu Ser Gln
               275                 280                 285
His Gly Val Phe Ser Glu Lys Glu Met Arg Phe Tyr Ala Thr Glu
               290                 295                 300
Ile Ile Leu Gly Leu Glu His Met His Asn Arg Phe Val Val Tyr
               305                 310                 315
Arg Asp Leu Lys Pro Ala Asn Ile Leu Leu Asp Glu His Gly His
               320                 325                 330
Ala Arg Ile Ser Asp Leu Gly Leu Ala Cys Asp Phe Ser Lys Lys
               335                 340                 345
```

PF-1506 P

```
Lys Pro His Ala Ser Val Gly Thr His Gly Tyr Met Ala Pro Glu
                350                 355                 360
Val Leu Gln Lys Gly Thr Ala Tyr Asp Ser Ser Ala Asp Trp Phe
                365                 370                 375
Ser Leu Gly Cys Met Leu Phe Lys Leu Leu Arg Gly His Ser Pro
                380                 385                 390
Phe Arg Gln His Lys Thr Lys Asp Lys His Glu Ile Asp Arg Met
                395                 400                 405
Thr Leu Thr Val Asn Val Glu Leu Pro Asp Thr Phe Ser Pro Glu
                410                 415                 420
Leu Lys Ser Leu Leu Glu Gly Leu Leu Gln Arg Asp Val Ser Lys
                425                 430                 435
Arg Leu Gly Cys His Gly Gly Gly Ser Gln Glu Val Lys Glu His
                440                 445                 450
Ser Phe Phe Lys Gly Val Asp Trp Gln His Val Tyr Leu Gln Lys
                455                 460                 465
Tyr Pro Pro Pro Leu Ile Pro Pro Arg Gly Glu Val Asn Ala Ala
                470                 475                 480
Asp Ala Phe Asp Ile Gly Ser Phe Asp Glu Glu Asp Thr Lys Gly
                485                 490                 495
Ile Lys Leu Leu Asp Cys Asp Gln Glu Leu Tyr Lys Asn Phe Pro
                500                 505                 510
Leu Val Ile Ser Glu Arg Trp Gln Gln Glu Val Thr Glu Thr Val
                515                 520                 525
Tyr Glu Ala Val Asn Ala Asp Thr Asp Lys Ile Glu Ala Arg Lys
                530                 535                 540
Arg Ala Lys Asn Lys Gln Leu Gly His Glu Glu Asp Tyr Ala Leu
                545                 550                 555
Gly Lys Asp Cys Ile Met His Gly Tyr Met Leu Lys Leu Gly Asn
                560                 565                 570
Pro Phe Leu Thr Gln Trp Gln Arg Arg Tyr Phe Tyr Leu Phe Pro
                575                 580                 585
Asn Arg Leu Glu Trp Arg Gly Glu Gly Glu Ser Arg Ser Asp Pro
                590                 595                 600
Glu Phe Val Gln Trp Lys Lys Glu Leu Asn Glu Thr Phe Lys Glu
                605                 610                 615
Ala Arg Arg Leu Leu Arg Arg Ala Pro Lys Phe Leu Asn Lys Pro
                620                 625                 630
Arg Ser Gly Thr Val Glu Leu Pro Lys Pro Ser Leu Cys His Arg
                635                 640                 645
Asn Ser Asn Gly Leu
                650
```

<210> 14
<211> 603
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7517238CD1

<400> 14

```
Met Lys Asp Tyr Asp Glu Leu Leu Lys Tyr Tyr Glu Leu His Glu
  1               5                  10                  15
Thr Ile Gly Thr Gly Gly Phe Ala Lys Val Lys Leu Ala Cys His
             20                  25                  30
Ile Leu Thr Gly Glu Met Val Ala Ile Lys Ile Met Asp Lys Asn
             35                  40                  45
Thr Leu Gly Ser Asp Leu Pro Arg Ile Lys Thr Glu Ile Glu Ala
             50                  55                  60
Leu Lys Asn Leu Arg His Gln His Ile Cys Gln Leu Tyr His Val
             65                  70                  75
Leu Glu Thr Ala Asn Lys Ile Phe Met Val Leu Glu Glu Asn Leu
             80                  85                  90
Leu Phe Asp Glu Tyr His Lys Leu Lys Leu Ile Asp Phe Gly Leu
```

95 100 105
Cys Ala Lys Pro Lys Gly Asn Lys Asp Tyr His Leu Gln Thr Cys
110 115 120
Cys Gly Ser Leu Ala Tyr Ala Ala Pro Glu Leu Ile Gln Gly Lys
125 130 135
Ser Tyr Leu Gly Ser Glu Ala Asp Val Trp Ser Met Gly Ile Leu
140 145 150
Leu Tyr Val Leu Met Cys Gly Phe Leu Pro Phe Asp Asp Asp Asn
155 160 165
Val Met Ala Leu Tyr Lys Lys Ile Met Arg Gly Lys Tyr Asp Val
170 175 180
Pro Lys Trp Leu Ser Pro Ser Ser Ile Leu Leu Leu Gln Gln Met
185 190 195
Leu Gln Val Asp Pro Lys Lys Arg Ile Ser Met Lys Asn Leu Leu
200 205 210
Asn His Pro Trp Ile Met Gln Asp Tyr Asn Tyr Pro Val Glu Trp
215 220 225
Gln Ser Lys Asn Pro Phe Ile His Leu Asp Asp Asp Cys Val Thr
230 235 240
Glu Leu Ser Val His His Arg Asn Asn Arg Gln Thr Met Glu Asp
245 250 255
Leu Ile Ser Leu Trp Gln Tyr Asp His Leu Thr Ala Thr Tyr Leu
260 265 270
Leu Leu Leu Ala Lys Lys Ala Arg Gly Lys Pro Val Arg Leu Arg
275 280 285
Leu Ser Ser Phe Ser Cys Gly Gln Ala Ser Ala Thr Pro Phe Thr
290 295 300
Asp Ile Lys Ser Asn Asn Trp Ser Leu Glu Asp Val Thr Ala Ser
305 310 315
Asp Lys Asn Tyr Val Ala Gly Leu Ile Asp Tyr Asp Trp Cys Glu
320 325 330
Asp Asp Leu Ser Thr Gly Ala Ala Thr Pro Arg Thr Ser Gln Phe
335 340 345
Thr Lys Tyr Trp Thr Glu Ser Asn Gly Ala Glu Ser Lys Ser Leu
350 355 360
Thr Pro Ala Leu Cys Arg Thr Pro Ala Asn Lys Leu Lys Asn Lys
365 370 375
Glu Asn Val Tyr Thr Pro Lys Ser Ala Val Lys Asn Glu Glu Tyr
380 385 390
Phe Met Phe Pro Glu Pro Lys Thr Pro Val Asn Lys Asn Gln His
395 400 405
Lys Arg Glu Ile Leu Thr Thr Pro Asn Arg Tyr Thr Thr Pro Ser
410 415 420
Lys Ala Arg Asn Gln Cys Leu Lys Glu Thr Pro Ile Lys Ile Pro
425 430 435
Val Asn Ser Thr Gly Thr Asp Lys Leu Met Thr Gly Val Ile Ser
440 445 450
Pro Glu Arg Arg Cys Arg Ser Val Glu Leu Asp Leu Asn Gln Ala
455 460 465
His Met Glu Glu Thr Pro Lys Arg Lys Gly Ala Lys Val Phe Gly
470 475 480
Ser Leu Glu Arg Gly Leu Asp Lys Val Ile Thr Val Leu Thr Arg
485 490 495
Ser Lys Arg Lys Gly Ser Ala Arg Asp Gly Pro Arg Arg Leu Lys
500 505 510
Leu His Tyr Asn Val Thr Thr Thr Arg Leu Val Asn Pro Asp Gln
515 520 525
Leu Leu Asn Glu Ile Met Ser Ile Leu Pro Lys Lys His Val Asp
530 535 540
Phe Val Gln Lys Gly Tyr Thr Leu Lys Cys Gln Thr Gln Ser Asp
545 550 555
Phe Gly Lys Val Thr Met Gln Phe Glu Leu Glu Val Cys Gln Leu
560 565 570
Gln Lys Pro Asp Val Val Gly Ile Arg Arg Gln Arg Leu Lys Gly
575 580 585
Asp Ala Trp Val Tyr Lys Arg Leu Val Glu Asp Ile Leu Ser Ser

PF-1506 P

590 595 600
Cys Lys Val

<210> 15
<211> 1916
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7517831CB1

<400> 15
gttaggccag gaggaccatg tgaatggggc cagagggctc ccgggctggg caggaccatg 60
ggctgtggct gcagctcaca cccggaagat gactggatgg aaaacatcga tgtgtgtgag 120
aactgccatt atcccatagt cccactggat ggcaagggca cgctgctcat ccgaaatggc 180
tctgagacaa cctggttatc gctctgcaca gctatgagcc ctctcacgac ggagatctgg 240
gctttgagaa gggggaacag ctccgcatcc tggagcagag cggcgagtgg tggaaggcgc 300
agtccctgac cacgggccag gaaggcttca tcccccttcaa ttttgtggcc aaagcgaaca 360
gcctggagcc cgaaccctgg ttcttcaaga acctgagccg caaggacgcg gagcggcagc 420
tcctggcgcc cgggaacact cacggctcct tcctcatccg ggagagcgag agcaccgcgg 480
gatcgttttc actgtcggtc cgggacttcg accagaacca gggagaggtg gtgaaacatt 540
acaagatccg taatctggac aacggtggct tctacatctc ccctcgaatc acttttcccg 600
gcctgcatga actggtccgc cattacacca atgcttcaga tgggctgtgc acacggttga 660
gccgcccctg ccagacccag aagccccaga agccgtggtg ggaggacgag tgggaggttc 720
ccagggagac gctgaagctg gtggagcggc tgggggctgg acagttcggg gaggtgtgga 780
tggggtacta caacgggcac acgaaggtgg cggtgaagag cctgaagcag ggcagcatgt 840
ccccggacgc cttcctggcc gaggccaacc tcatgaagca gctgcaacac cagcggctgg 900
ttcggctcta cgctgtggtc acccaggagc ccatctacat catcactgaa tacatggaga 960
atgggagtct agtggatttt ctcaagaccc cttcaggcat caagttgacc atcaacaaac 1020
tcctggacat ggcagcccaa ttgcagaagg catggcattc attgaagagc ggaattatat 1080
tcatcgtgac cttcgggctg ccaacattct ggtgtctgac accctgagct gcaagattgc 1140
agacttgggt ctagcacgcc tcattgagga caacgagtac acagccaggg agggggccaa 1200
gtttcccatt aagtggacag cgccagaagc cattaactac gggacattca ccatcaagtc 1260
agatgtgtgg tcttttggga tcctgctgac ggaaattgtc acccacggcc gcatcccttа 1320
cccagggatg accaacccgg aggtgattca gaacctggag cgaggctacc gcatggtgcg 1380
ccctgacaac tgtccagagg agctgtacca actcatgagg ctgtgctgga aggagcgccc 1440
agaggaccgg cccacctttg actacctgcg cagtgtgctg gaggacttct tcacggccac 1500
agagggccag taccagcctc agccttgaga ggccttgaga ggccctgggg ttctcccct 1560
ttctctccag cctgacttgg ggagatggag ttcttgtgcc atagtcacat ggcctatgca 1620
catatggact ctgcacatga atcccaccca catgtgacac atatgcacct tgtgtctgta 1680
cacgtgtcct gtagttgcgt ggactctgca catgtcttgt acatgtgtag cctgtgcatg 1740
tatgtcttgg acactgtaca aggtacccct ttctggctct cccatttcct gagaccacag 1800
agagagggga gaagcctggg attgacagaa gcttctgccc acctactttt ctttcctcag 1860
atcatccaga agttcctcaa gggccaggac tttatctaat acctctgtgt gctcct 1916

<210> 16
<211> 926
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7520272CB1

<400> 16
tctacccggt tcaagcatgg ctgaccaggc gcccttcgac acggacgtca acaccctgac 60
ccgcttcgtc atggaggagg gcaggaaggc ccgcggcacg ggcgagttga cccagctgct 120
caactcgctc tgcacagcag tcaaagccat ctcttcggcg gtgcgcaagg cgggcatcgc 180
gcacctctat ggcattgctg gttctaccaa cgtgacaggt gatcaagtta agaagctgga 240
cgtcctctcc aacgacctgg ttatgaacat gttaaagtca tcctttgcca cgtgtgttct 300
cgtgtcagaa gaagataaac acgccatcat agtggaaccg gagaaaaggg gtaaatatgt 360
ggtctgtttt gatcccctg atggatcttc caacatcgat tgccttgtgt ccgttggaac 420
catttttggc atctatagaa agaaatcaac tgatgagcct tctgagaagg atgctctgca 480
accaggccgg aacctggtgg cagccggcta cgcactgtat ggcagtgcca ccatgctggt 540

ccttgccatg gactgtgggg tcaactgctt catgctggac ccggataatt cagctcctta 600
tggggcccgg tatgtgggct ccatggtggc tgatgttcat cgcactctgg tctacggagg 660
gatatttctg taccccgcta acaagaagag ccccaatgga aagctgagac tgctgtacga 720
atgcaacccc atggcctacg tcatggagaa ggctggggga atggccacca ctgggaagga 780
ggccgtgtta gacgtcattc ccacagacat tcaccagagg gcgccggtga tcttggggtc 840
ccccgacgac gtgctcgagt tcctgaaggt gtatgagaag cactctgccc agtgagcacc 900
tgccctgcct gcatctggag aattga 926

<210> 17
<211> 1382
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7521279CB1

<400> 17
tgatggcgtc cccacgggaa ttgacacaga accccctgaa gaagatctgg atgccataca 60
gcaatgggcg gcccgctctg cacgcttgcc agcgcggtgt gtgcatgacc aactgcccaa 120
ctctcattgt catggtgggc ctgcccgcca ggggcaagac ctacatctcc aagaagctga 180
ctcgatacct gaactggatt ggtgtgccca ctcgggagtt caatgttggc cagtatcgcc 240
gggacgtggt caagacctac aaatcttttg aatttttct ccccgacaat gaagagggcc 300
tgaaaatcag gaagcagtgt gccctggcag ccctccgtga cgtccggcgg ttccttagtg 360
aggagggggg acatgtggcg gtttttgatg ccacaaacac cacccgagaa cggagagcga 420
ccatctttaa ttttggagaa cagaatggct acaagacctt ttttgtcgag tccatctgtg 480
tggatcctga ggtcatagct gccaacatcg tgcaagtgaa actgggcagc cctgactatg 540
tcaaccgcga cagtgatgag gctacggagg acttcatgag gcgcattgag tgctatgaga 600
actcctacga gtcgctagat gaggacctgg atagggacct gtcctatatc aagatcatgg 660
atgtgggcca gagctacgtg gtgaaccgtg tggctgacca catccagagc cgcatcgtat 720
attacctcat gaacatccac gtgacccccc gctccatcta cctctgccgg cacgggggaga 780
gcgagctcaa cctcaagggc cggattggcg gggacccagg actgtcccct cggggcaggg 840
agtttgccaa gagtctagcc cagttcatca gtgaccaaaa tatcaaggat ctgaaggtct 900
ggacaagcca gatgaagagg acaatccaga cggctgaggc actgggtgtg ccctatgaac 960
agtggaaggt cctcaacgag atcgatgcgt cctacgagga cctggtccag agactggagc 1020
ctgtcatcat ggagctggag aggcaagaga atgtgctggt catctgccac caggctgtga 1080
tgcgctgcct gctggcctac ttcctcgaca aggcagcaga acagctgccc tacctcaagt 1140
gtccgctgca cacagtcctg aagctgactc ctgtggcata tggttgtaaa gtggagtcca 1200
tattcctgaa cgtggctgct gtgaacacgc accgggacag gcctcagaac gtggacatct 1260
caagacctcc agaggaagcc cttgtcacgg tgcctgctca ccagtgacca tgttcatcca 1320
ctgtgaccac taggcaggca ctgctctctg cagagggggt cattccaggc cctccagtgt 1380
ga 1382

<210> 18
<211> 1678
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7523965CB1

<400> 18
tatggccgca ttgtaccgcc ctggcctgcg gcttaactgg catgggctga gccccttggg 60
ctggccatca tgccgtagca tccagaccct gcgagtgctt agtggagatc tgggccagct 120
tcccactggc attcgagatt ttgtagagca cagtgcccgc ctgtgccaac cagagggcat 180
ccacatctgt gatggaactg aggctgagaa tactgccaca ctgaccctgc tggagcagca 240
gggcctcatc cgaaagctcc ccaagtacaa taactgctgg ctggcccgca cagaccccaa 300
ggatgtggca cgagtagaga gcaagacggt gattgtaact ccttctcagc gggacacggt 360
accactcccg cctggtgggg cccgtgggca gctgggcaac tggatgtccc cagctgattt 420
ccagcgagct gtggatgaga ggtttccagg ctgcatgcag ggccgcacca tgtatgtgct 480
tccattcagc atgggtcctg tgggctcccc gctgtcccgc atcggggtgc agctcactga 540
ctcagcctat gtggtggcaa gcatgcgtat tatgacccga ctggggacac ctgtgcttca 600
ggccctggga gatggtgact ttgtcaagtg tctgcactcc gtgggccagc cctgacagg 660
acaagatcct gggcatcacc agccctgcag ggaagaagcg ctatgtggca gccgccttcc 720
ctagtgcctg tggcaagacc aacctggcta tgatgcggcc tgcactgcca ggctggaaag 780

PF-1506 P

tggagtgtgt gggggatgat attgcttgga tgaggtttga cagtgaaggt cgactccggg 840
ccatcaaccc tgagaacggc ttctttgggg ttgcccctgg tacctctgcc accaccaatc 900
ccaacgccat ggctacaatc cagagtaaca ctatttttac caatgtggct gagaccagtg 960
atggtggcgt gtactgggag ggcattgacc agcctcttcc acctggtgtt actgtgacct 1020
cctggctggg caaaccctgg aaacctggtg acaaggagcc ctgtgcacat cccaactctc 1080
gattttgtgc cccggctcgc cagtgcccca tcatggaccc agcctgggag gccccagagg 1140
gtgtccccat tgacgccatc atctttggtg gccgcagacc caaagggaag atcatcatgc 1200
acgacccatt tgccatgcgg ccctttttttg gctacaactt cgggcactac ctggaacact 1260
ggctgagcat ggaagggcgc aaggggggccc agctgccccg tatcttccat gtcaactggt 1320
tccggcgtga cgaggcaggg cacttcctgt ggccaggctt tggggagaat gctcgggtgc 1380
tagactggat ctgccggcgg ttagaggggg aggacagtgc ccgagagaca cccattgggc 1440
tggtgccaaa ggaaggagcc ttggatctca gcggcctcag agctatagac accactcagc 1500
tgttctccct ccccaaggac ttctgggaac aggaggttcg tgacattcgg agctacctga 1560
cagagcaggt caaccaggat ctgcccaaag aggtgttggc tgagcttgag gccctggaga 1620
gacgtgtgca caaaatgtga cctgaggccc tagtctagca agaggacata gcacccta 1678

<210> 19
<211> 895
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7524016CB1

<400> 19
tctgtaaatg caagagaacc gagtgtggat aattagcgat ggaagaaaaa acctctagaa 60
taaaagcatc catacccccag tttaccaatt cccccacaat ggtgatcatg gtgggtttac 120
cagctcgagg caagacctat atctccacaa agctcacacg atatctcaac tggataggaa 180
caccaactaa agtgtttaat ttaggccagt atcgacgaga ggcagtgagc tacaagaact 240
atgaattctt tcttccagac aacatggaag ccctgcaaat caggaagcag tgcgccctgg 300
cagccctgaa ggatgttcac aactatctca gccatgagga aggtcatgtt gcggtttttg 360
atgccaccaa cactaccaga gaacgacggt cactgatcct gcagtttgca aaagaacatg 420
gttacaaggt gtttttcatt gagtccattt gtaatgaccc tggcataatt gcagaaaaca 480
tcaggcaagt gaaacttggc agccctgatt atatagactg tgaccgggaa aaggttctgg 540
aagactttct aaagagaatt gagtgctatg aggtcaacta ccaacccttg gatgaggaac 600
tggacagatc ttcgacgtgg gcacacgcta catggtgaac cgagtgcagg atcacatcca 660
gagccgcaca gtctactacc tcatgaatat ccatgtcaca cctcgctcca tctacctttg 720
ccgacatggc gagagtgaac tcaacatcag aggccgcatc ggaggtgact ctggcctctc 780
agttcgcggc aagcagtatg cctatgccct ggccaacttc attcagtccc agggcatcag 840
ctccctgaag gtgtggacca gtcacatgaa gaggaccatc cagacagctg aggcc 895

<210> 20
<211> 1294
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7524680CB1

<400> 20
tctgtaaatg caagagaacc gagtgttgga taattagcga tggaagaaaa aacctctaga 60
ataaaagtgt ttaatttagg ccagtatcga cgagaggcag tgagctacaa gaactatgaa 120
ttctttcttc cagacaacat ggaagccctg caaatcagga agcagtgcgc cctggcagcc 180
ctgaaggatg ttcacaacta tctcagccat gaggaaggtc atgttgcggt ttttgatgcc 240
accaacacta ccagagaacg acggtcactg atcctgcagt ttgcaaaaga acatggttac 300
aaggtgtttt tcattgagtc catttgtaat gaccctggca taattgcaga aaacatcagg 360
caagtgaaac ttggcagccc tgattatata gactgtgacc gggaaaaggt tctggaagac 420
tttctaaaga gaattgagtg ctatgaggtc aactaccaac ccttggatga ggaactggac 480
agccacctgt cctacatcaa gatcttcgac gtgggcacac gctacatggt gaaccgagtg 540
caggatcaca tccagagccg cacagtctac tacctcatga atatccatgt cacacctcgc 600
tccatctacc tttgccgaca tggcgagagt gaactcaaca tcagaggccg catcggaggt 660
gactctggcc tctcagttcg cggcaagcag tatgcctatg ccctggccaa cttcattcag 720
tcccagggca tcagctccct gaaggtgtgg accagtcaca tgaagaggac catccagaca 780
gctgaggccc tgggtgtccc ctatgagcag tggaaggccc tgaatgagat tgatgcgggt 840

PF-1506 P

gtctgtgagg agatgaccta tgaagaaatc caggaacatt accctgaaga atttgcactg 900
cgagaccaag ataaatatcg ctaccgctat cccaagggag agtcctatga ggatctggtt 960
cagcgtctgg agccagtgat aatggagcta gaacgacagg agaatgtact ggtgatctgc 1020
caccaggctg tcatgcggtg cctcctggcc tatttcctgg ataaaagttc agatgagctt 1080
ccatatctca agtgccctct gcacacagtg ctcaaactca ctcctgtggc ttatggctgc 1140
aaagtggaat ccatctacct gaatgtggag accgtgaaca cacaccggga gaagcctgag 1200
aatgtggaca tcacccggga acctgaggaa gccctggata ctgtcccagc ccactactga 1260
gccctttcca agaagtcaaa ctgcctgtgt ccta 1294

<210> 21
<211> 1354
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7524757CB1

<400> 21
tctgtaaatg caagagaacc gagtgttgga taattagcga tggaagaaaa aacctctaga 60
ataaaagcat ccatacccca gtttaccaat tcccccacaa tggtgatcat ggtgggttta 120
ccagctcgag gcaagaccta tatctccaca aagctcacac gatatctcaa ctggatagga 180
acaccaacta aagacaacat ggaagccctg caaatcagga agcagtgcgc cctggcagcc 240
ctgaaggatg ttcacaacta tctcagccat gaggaaggtc atgttgcggt ttttgatgcc 300
accaacacta ccagagaacg acggtcactg atcctgcagt ttgcaaaaga acatggttac 360
aaggtgtttt tcattgagtc catttgtaat gaccctggca taattgcaga aaacatcagg 420
caagtgaaac ttggcagccc tgattatata gactgtgacc gggaaaaggt tctggaagac 480
tttctaaaga gaattgagtg ctatgaggtc aactaccaac ccttggatga ggaactggac 540
agccacctgt cctacatcaa gatcttcgac gtgggcacac gctacatggt gaaccgagtg 600
caggatcaca tccagagccg cacagtctac taccttcatga atatccatgt cacacctcgc 660
tccatctacc tttgccgaca tggcgagagt gaactcaaca tcagaggccg catcggaggt 720
gactctggcc tctcagttcg cggcaagcag tatgcctatg ccctggccaa cttcattcag 780
tcccagggca tcagctccct gaaggtgtgg accagtcaca tgaagaggac catccagaca 840
gctgaggccc tgggtgtccc ctatgagcag tggaaggccc tgaatgagat tgatgcgggt 900
gtctgtgagg agatgaccta tgaagaaatc cgggaacatt accctgaaga atttgcactg 960
cgagaccaag ataaatatcg ctaccgctat cccaagggag agtcctatga ggatctggtt 1020
cagcgtctgg agccagtgat aatggagcta gaacgacagg agaatgtact ggtgatctgc 1080
caccaggctg tcatgcggtg cctcctggcc tatttcctgg ataaaagttc agatgagctt 1140
ccatatctca agtgccctct gcacacagtg ctcaaactca ctcctgtggc ttatggctgc 1200
aaagtggaat ccatctacct gaatgtggag gccgtgaaca cacaccggga gaagcctgag 1260
aatgtggaca tcacccggga acctgaggaa gccctggata ctgtcccagc ccactactga 1320
gccctttcca agaagtcaaa ctgcctgtgt ccta 1354

<210> 22
<211> 1204
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7516229CB1

<400> 22
agagggcggg tccccggcct cgggagcacg gcggtggagg ggacatagga ggcggccatg 60
gcgacccccg gcaacctagg gtcctctgtc ctggcgagca agaccaagac caagaagaag 120
cacttcgtag cgcagaaagt gaagctgttt cgggccagcg acccgctgct cagcgtcctc 180
atgtgggggg taaaccactc gatcaatgaa ctgagccatg ttcaaatccc tgttatgttg 240
atgccagatg acttcaaagc ctattcaaaa ataaaggtgg acaatcacct ttttaacaaa 300
gaaaacatgc cgagccattt caagtttaag gaatactgcc cgatggtctt ccgtaacctg 360
cgggagaggt ttggaattga tgatcaagat ttccagtaca tagtggaatg tcatgggatc 420
acccttcttc cccagttctt gggcatgtac cggcttaatg ttgatggagt tgaaatatat 480
gtgatagtta caagaaatgt attcagccac cgtttgtctg tgtataggaa atacgactta 540
aagggctcta cagtggctag agaagctagt gacaaagaaa aggccaaaga actgccaact 600
ctgaaagata atgatttcat taatgagggc caaaagattt atattgatga caacaacaag 660
aaggtcttcc tggaaaaact aaaaaaggat gttgagtttc tggcccagct gaagctcatg 720
gactacagtc tgctggtggg aattcatgat gtggagagag ccgaacagga ggaagtggag 780

PF-1506 P

tgtgaggaga acgatgggga ggaggagggc gagagcgatg gcacccaccc ggtgggaacc 840
cccccagata gccccgggaa tacactgaac agctcaccac ccctggctcc cggggagttc 900
gatccgaaca tcgacgtcta tggaattaag tgccatgaaa actcgcctag gaaggaggtg 960
tacttcatgg caattattga catccttact cattatgatg caaaaaagaa agctgcccat 1020
gctgcaaaaa ctgttaaaca tggcgctggc gcggagatct ccaccgtgaa cccagaacag 1080
tattcaaagc gctttttgga ctttattggc cacatcttga cgtaacctcc tgcgcagcct 1140
cggacagaca tgaacattgg agggacagag gtggcttcga agaacgaatt gcagcacact 1200
gcgg 1204

<210> 23
<211> 1859
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7516525CB1

<400> 23
caacaaggct acgcagaaga acccccttga ctgaagcaat ggaggggggt ccagctgtct 60
gctgccagga tcctcgggca gagctggtag aacgggtggc agccatcgat gtgactcact 120
tggaggaggc agatggtggc ccagagccta ctagaaacgg tgtggacccc ccaccacggg 180
ccagagctgc ctctgtgatc cctggcagta cttcaagact gctcccagcc cggcctagcc 240
tctcagccag gaagctttcc ctacaggagc ggccagcagg aagctatctg gaggcgcagg 300
ctgggcctta tgccacgggg cctgccagcc acatctcccc ccgggcctgg cggaggccca 360
ccatcgagtc ccaccacgtg gccatctcag atgcagagga ctgcgtgcag ctgaaccagt 420
acaagctgca gagtgagatt ggcaagggtg cctacggtgt ggtgaggccg gcctacaacg 480
aaagtgaaga cagacactat gcaatgaaag tcctttccaa aaagaagtta ctgaagcagt 540
atggctttcc acgtcgccct cccccgagag ggtcccaggc tgcccaggga ggaccagcca 600
agcagctgct gcccctggag cgggtgtacc aggagattgc catcctgaag aagctggacc 660
acgtgaatgt ggtcaaactg atcgaggtcc tggatgaccc agctgaggac aacctctatt 720
tggccctgca gaaccaggcc cagaatatcc agttagattc aacaaatatc gccaagcccc 780
actccctgct tccctctgag cagcaagaca gtggatccac gtgggctgcg cgctcagtgt 840
ttgacctcct gagaaagggg cccgtcatgg aagtgccctg tgacaagccc ttctcggagg 900
agcaagctcg cctctacctg cgggacgtca tcctgggcct cgagtacttg cactgccaga 960
agatcgtcca cagggacatc aagccatcca acctgctcct gggggatgat gggcacgtga 1020
agatcgccga ctttggcgtc agcaaccagt ttgaggggaa cgacgctcag ctgtccagca 1080
cggcgggaac cccagcattc atggcccccg aggccatttc tgattccggc cagagcttca 1140
gtgggaaggc cttggatgta tgggccactg gcgtcacgtt gtactgcttt gtctatggga 1200
agtgcccgtt catcgacgat ttcatcctgg ccctccacag gaagatcaag aatgagcccg 1260
tggtgtttcc tgaggggcca gaaatcagcg aggagctcaa ggacctgatc ctgaagatgt 1320
tagacaagaa tcccgagacg agaattgggg tgccagacat caagttgcac ccttgggtga 1380
ccaagaacgg ggaggagccc attccttcgg aggaggagca ctgcagcgtg gtggaggtga 1440
cagaggagga ggttaagaac tcagtcaggc tcatccccag ctggaccacg gtgatcctgg 1500
tgaagtccat gctgaggaag cgttcctttg ggaacccgtt tgagccccaa gcacggaggg 1560
aagagcgatc catgtctgct ccaggaaacc tactggtgaa agaagggttt ggtgaagggg 1620
gcaagagccc agagctcccc ggcgtccagg aagacgaggc tgcatcctga gcccctgcat 1680
gcacccaggg ccacccggca gcacactcat cccgcgcctc cagaggccca cccctcatgc 1740
aacagccgcc cccgcaggca gggggctggg gactgcagcc ccactcccgc ccctccccca 1800
tcgtgctgca tgacctccac gcacgcacgt ccagggacag actggaatgt atgtcattt 1859

<210> 24
<211> 1695
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7516533CB1

<400> 24
actgagccta agcagccggt gatggcggca gcggctgtgg tggctgcggc gggtccgggc 60
ccatgaggcg acgaaggagg cgggacggct tttacccagc cccggacttc cgagacaggg 120
aagctgagga catggcagga gtgtttgaca tagacctgga ccagccagag gacgcgggct 180
ctgaggatga gctggaggag ggggcaatga tagtaagaaa tgctaaagat acagctcata 240
caaaagcaga acggaatatt ctggaggaag taaagcatcc cttcatcgtg gatttaattt 300

PF-1506 P

atgcctttca gactggtgga aaactctacc tcatccttga gtatctcagt ggaggagaac 360
tatttatgca gttagaaaga gagggaatat ttatggaaga cactgcctgc ttttacttgg 420
cagaaatctc catggctttg gggcatttac atcaaaaggg gatcatctac agagacctga 480
agccggagaa tatcatgctt aatcaccaag gtcatgtgaa actaacagac tttggactat 540
gcaaagaatc tattcatgat ggaacagtca cacacacatt ttgtggaaca atagaataca 600
tggcccctga aatcttgatg agaagtggcc acaatcgtgc tgtggattgg tggagtttgg 660
gagcattaat gtatgacatg ctgactggag cacccccatt cactggggag aatagaaaga 720
aaacaattga caaaatcctc aaatgtaaac tcaatttgcc tccctacctc acacaagaag 780
ccagagatct gcttaaaaag ctgctgaaaa gaaatgctgc ttctcgtctg ggagctggtc 840
ctggggacgc tggagaagtt caagctcatc cattctttag acacattaac tgggaagaac 900
ttctggctcg aaaggtggag ccccccttta aacctctgtt gcaatctgaa gaggatgtaa 960
gtcagtttga ttccaagttt acacgtcaga cacctgtcga cagcccagat gactcaactc 1020
tcagtgaaag tgccaatcag gtctttctgg gttttacata tgtggctcca tctgtacttg 1080
aaagtgtgaa agaaaagttt tcctttgaac caaaaatccg atcacctcga agatttattg 1140
gcagcccacg aacacctgtc agcccagtca aattttctcc tggggatttc tggggaagag 1200
gtgcttcggc cagcgcagca aatcctcaga cacctgtgga atacccaatg gaaacaagtg 1260
gcatagagca gatggatgtg acaatgagtg gggaagcatc ggcaccactt ccaatacgac 1320
agccgaactc tgggccatac aaaaaacaag cttttcccat gatctccaaa cggccagagc 1380
acctgcgtat gaatctatga cagagcatgc ttttaatgaa tttaaggcaa aaaggtggag 1440
agggagatgt gtgagcatcc tgcaaggtga aacgactcaa aatgacagtt tcagagagtc 1500
aatgtcatta catagaacac ttcagacaca ggaaaaataa acgtggattt taaaaaatca 1560
atcaatggtg caaaaaaaaa cttaaagcaa atagtattgc tgaactctta ggcacatcaa 1620
ttaattgatt cctcgcgaca tcttctcaac cttatcaagg attttcatgt tgatgactcg 1680
aaactgacag tatta 1695

<210> 25
<211> 3891
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7516613CB1

<400> 25
agagcgacag agacatttat tgttatttgt tttttggtgg caaaaaggga aaatggcgaa 60
cgactcccct gcaaaaagtc tggtggacat cgacctctcc tccctgcggg atcctgctgg 120
gatttttgag ctggtggaag tggttggaaa tggcacctat ggacaagtct ataagggtcg 180
acatgttaaa acgggtcagt tggcagccat caaagttatg gatgtcactg aggatgaaga 240
ggaagaaatc aaactggaga taaatatgct aaagaaatac tctcatcaca gaaacattgc 300
aacatattat ggtgctttca tcaaaaagag ccctccagga catgatgacc aactctggct 360
tgttatggag ttctgtgggg ctgggtccat tacagacctt gtgaagaaca ccaaagggaa 420
cacactcaaa gaagactgga tcgcttacat ctccagagaa atcctgaggg gactggcaca 480
tcttcacatt catcatgtga ttcaccggga tatcaagggc cagaatgtgt tgctgactga 540
gaatgcagag gtgaaacttg ttgactttgg tgtgagtgct cagctggacg ggactgtggg 600
gcggagaaat acgttcatag gcactcccta ctggatggct cctgaggtca tcgcctgtga 660
tgagaaccca gatgccacct atgattacag aagtgatctt tggtcttgtg gcattacagc 720
cattgagatg ggggaaggtg ctccccctct ctgtgacatg catccaatga gagcactgtt 780
tctcattccc agaaaccctc ctccccggct gaagtcaaaa aaatggtcga agaagttttt 840
tagttttata gaagggtgcc tggtgaagaa ttacatgcag cggccctcta cagagcagct 900
tttgaaacat ccttttataa gggatcagcc aaatgaaagg caagttagaa tccagcttaa 960
ggatcatata gatcgtacca ggaagaagag aggcgagaaa gatgaaactg agtatgagta 1020
cagtgggagt gaggaagaag aggaggaagt gcctgaacag gaaggagagc caagttccat 1080
tgtgaacgtg cctggtgagt ctactcttcg ccgagatttc ctgagactgc agcaggagaa 1140
caaggaacgt tccgaggctc ttcggagaca acagttacta caggagcaac agctccggga 1200
gcaggaagaa tataaaaggc aactgctggc agagagacag aagcggattg agcagcagaa 1260
agaacagagg cgacggctag aagagcaaca aaggagagag cgggaagcta gaaggcagca 1320
ggaacgtgaa cagcgaagga gagaacaaga agaaaagagg cgtctagagg agttggagag 1380
aaggcgcaaa gaagaagagg agaggagaca ggcagaagaa gaaaagagga gagttgaaag 1440
agaacaggag tatatcaggc gacagctaga agaggagcag cggcacttgg aagtccttca 1500
gcagcagctg ctccaggagc aggccatgtt actgcatgac cataggaggc cgcaccgcga 1560
gcactcgcag cagccgccac caccgcagca ggaaaggagc aagccaagct tccatgctcc 1620
cgagcccaaa gcccactacg agcctgctga ccgagcgcga gaggttcctg tgagaacaac 1680
atctcgctcc cctgttctgt cccgtcgaga ttccccactg cagggcagtg ggcagcagaa 1740
tagccaggca ggacagagaa actccaccag cagtattgag cccaggcttc tgtgggagag 1800
agtggagaag ctgatgccca gacctggcag tggcagctcc tcagggtcca gcaactcagg 1860

atcccagccc gggtctcacc ctgggtctca gagtggctcc ggggaacgct tcagagtgag 1920
atcatcatcc aagtctgaag gctctccatc tcagcgcctg gaaaatgcag tgaaaaaacc 1980
tgaagataaa aaggaagttt tcagacccct caagcctgct gatctgaccg cactggccaa 2040
agagcttcga gcagtggaag atgtacggcc acctcacaaa gtaacggact actcctcatc 2100
cagtgaggag ccggggacga cggatgagga ggacgacgat gtggagcagg aaggggctga 2160
cgagtccacc tcaggaccag aggacaccag agcagcgtca tctctgaatt tgagcaatgg 2220
tgaaacggaa tctgtgaaaa ccatgattgt ccatgatgat gtagaaagtg agccggccat 2280
gaccccatcc aaggagggca ctctaatcgt ccgccagagt acagttgacc aaaagcgtgc 2340
cagccatcat gagagcaatg gctttgccgg tcgcattcac ctcttgccag atctcttaca 2400
gcaaagccat tcctcctcca cttcctccac ctcctcctcc ccatcctcca gccagccgac 2460
acccaccatg tccccacaga caccccagga caagctcact gctaatgaga ctcagtccgc 2520
tagtagcaca ctccagaaac acaaatcttc ctcctccttt acacctttta tagaccccag 2580
attactacag atttctccat ctagcggaac aacagtgaca tctgtggtgg gattttcctg 2640
tgatgggatg agaccagaag ccataaggca agatcctacc cggaaaggct cagtggtcaa 2700
tgtgaatcct accaacacta ggccacagag tgacaccccg gagattcgta aatacaagaa 2760
gaggtttaac tctgagattc tgtgtgctgc cttatgggga gtgaatttgc tagtgggtac 2820
agagagtggc ctgatgctgc tggacagaag tggccaaggg aaggtctatc ctcttatcaa 2880
ccgaagacga tttcaacaaa tggacgtact tgagggcttg aatgtcttgg tgacaatatc 2940
tggcaaaaag gataagttac gtgtctacta tttgtcctgg ttaagaaata aaatacttca 3000
caatgatcca gaagttgaga agaagcaggg atggacaacc gtaggggatt tggaaggatg 3060
tgtacattat aaagttgtaa aatatgaaag aatcaaattt ctggtgattg ctttgaagag 3120
ttctgtggaa gtctatgcgt gggcaccaaa gccatatcac aaatttatgg cctttaagtc 3180
atttggagaa ttggtacata agccattact ggtggatctc actgttgagg aaggccagag 3240
gttgaaagtg atctatggat cctgtgctgg attccatgct gttgatgtgg attcaggatc 3300
agtctatgac atttatctac caacacatat ccagtgtagc atcaaacccc atgcaatcat 3360
catcctcccc aatacagatg gaatggagct tctggtgtgc tatgaagatg agggggttta 3420
tgtaaacaca tatggaagga tcaccaagga tgtagttcta cagtggggag agatgcctac 3480
atcagtagca tatattcgat ccaatcagac aatgggctgg ggagagaagg ccatagagat 3540
ccgatctgtg gaaactggtc acttggatgg tgtgttcatg cacaaaaggg ctcaaagact 3600
aaaattcttg tgtgaacgca atgacaaggt gttctttgcc tctgttcggt ctggtggcag 3660
cagtcaggtt tatttcatga ccttaggcag gacttctctt ctgagctggt agaagcagtg 3720
tgatccaggg attactggcc tccagagtct tcaagatcct gagaacttgg aattccttgt 3780
aactggagct cggagctgca ccgagggcaa ccaggacagc tgtgtgtgca gacctcatgt 3840
gttgggttct ctcccctcct tcctgttcct cttatatacc agtttatccc c 3891

<210> 26
<211> 3954
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7517068CB1

<400> 26
atttctccag cgaagaagta gacatggcga gcgactcccc ggctcgaagc ctggatgaaa 60
tagatctctc ggctctgagg gaccctgcag ggatctttga attggtggaa cttgttggaa 120
atggaacata cgggcaagtt tataagggtc gtcatgtcaa aacgggccag cttgcagcca 180
tcaaggttat ggatgtcaca ggggatgaag aggaagaaat caaacaagaa attaacatgt 240
tgaagaaata ttctcatcac cggaatattg ctacatacta tggtgctttt atcaaaaaga 300
acccaccagg catggatgac caactttggt tggtgatgga gttttgtggt gctggctctg 360
tcaccgacct gatcaagaac acaaaaggta acacgttgaa agaggagtgg attgcataca 420
tctgcaggga aatcttacgg gggctgagtc acctgcacca gcataaagtg attcatcgag 480
atattaaagg gcaaaatgtc ttgctgactg aaaatgcaga agttaaacta gtggactttg 540
gagtcagtgc tcagcttgat cgaacagtgg gcaggaggaa tactttcatt ggaactccct 600
actggatggc accagaagtt attgcctgtg atgaaaaccc agatgccaca tatgatttca 660
agagtgactt gtggtctttg ggtatcaccg ccattgaaat ggcagaaggt gctcccccctc 720
tctgtgacat gcaccccatg agagctctct tcctcatccc ccggaatcca gcgcctcggc 780
tgaagtctaa gaagtggtca aaaaaattcc agtcatttat tgagagctgc ttggtaaaga 840
atcacagcca gcgaccagca acagaacaat tgatgaagca tccatttata cgagaccaac 900
ctaatgagcg acaggtccgc attcaactca aggaccatat tgatagaaca aagaagaagc 960
gaggagaaaa agatgagaca gagtatgagt acagtggaag tgaggaagaa gaggaggaga 1020
atgactcagg agagcccagc tccatcctga atctgccagg ggagtcgacg ctgcggaggg 1080
actttctgag gctgcagctg gccaacaagg agcgttctga ggccctacgg aggcagcagc 1140
tggagcagca gcagcgggag aatgaggagc acaagcggca gctgctggcc gagcgtcaga 1200
agcgcatcga ggagcagaaa gagcagaggc ggcggctgga ggagatccca catctggtag 1260

PF-1506 P

ctgtaaaatc ccagggacct gccttgaccg cctcccagtc agtgcacgag cagcccacaa 1320
agggcctctc tgggtttcag gaggctctga acgtgacctc ccaccgcgtg gagatgccac 1380
gccagaactc agatcccacc tcggaaaatc ctcctctccc cactcgcatt gaaaagtttg 1440
accgaagctc ttggttacga caggaagaag acattccacc aaaggtgcct caaagaacaa 1500
cttctatatc cccagcatta gccagaaaga attctcctgg gaatggtagt gctctgggac 1560
ccagactagg atctcaaccc atcagagcaa gcaaccctga tctccggaga actgagccca 1620
tcttggagag ccccttgcag aggaccagca gtggcagttc ctccagctcc agcaccccta 1680
gctcccagcc cagctcccaa ggaggctccc agcctggatc acaagcagga tccagtggac 1740
gcaccagagt tcgagccaac agtaagtcag aaggatcacc tgtgcttccc catgagcctg 1800
ccaaggtgaa accagaagaa tccagggaca ttacccggcc cagtcgacca gctgatctga 1860
cggcattagc caaagaacta agagaactcc ggattgaaga aacaaaccgc ccaatgaaga 1920
aggtgactga ttactcctcc tccagtgagg agtcagaaag tagcgaggaa gaggaggaag 1980
atggagagag cgagacccat gatgggacag tggctgtcag cgacataccc agactgatac 2040
caacaggagc tccaggcagc aacgagcagt acaatgtggg aatggtgggg acgcatgggc 2100
tggagacctc tcatgcggac agtttcagcg gcagtatttc aagagaagga accttgatga 2160
ttagagagac gtctggagag aagaagcgat ctggccacag tgacagcaat ggctttgctg 2220
gccacatcaa cctccctgac ctggtgcagc agagccattc tccagctgga accccgactg 2280
agggactggg gcgcgtctca acccattccc aggagatgga ctctgggact gaatatggca 2340
tggggagcag caccaaagcc tccttcaccc cctttgtgga ccccagagta taccagacgt 2400
ctcccactga tgaagatgaa gaggatgagg aatcatcagc cgcagctctg tttactagcg 2460
aacttcttag gcaagaacag gccaaactca atgaagcaag aaagatttcg gtggtaaatg 2520
taaacccaac caacattcgg cctcatagcg acacaccaga aatcagacaa tacaagaaac 2580
gattcaactc agaaatactt tgtgcagctc tgtggggtgt aaaccttctg gtggggactg 2640
aaaatggcct gatgcttttg gaccgaagtg ggcaaggcaa agtctataat ctgatcaacc 2700
ggaggcgatt tcagcagatg gatgtgctag agggactgaa tgtccttgtg acaatttcag 2760
gaaagaagaa taagctacga gtttactatc tttcatggtt aagaaacaga atactacata 2820
atgacccaga agtagaaaag aaacaaggct ggatcactgt tggggacttg gaaggctgta 2880
tacattataa agttgttaaa tatgaaagga tcaaattttt ggtgattgcc ttaaagaatg 2940
ctgtggaaat atatgcttgg gctcctaaac cgtatcataa attcatggca tttaagtctt 3000
ttgcagatct ccagcacaag cctctgctag ttgatctcac ggtagaagaa ggtcaaagat 3060
taaaggttat ttttggttca cacactggtt tccatgtaat tgatgttgat tcaggaaact 3120
cttatgatat ctacatacca tctcatattc agggcaatat cactcctcat gctattgtca 3180
tcttgcctaa aacagatgga atggaaatgc ttgtttgcta tgaggatgag ggggtttatg 3240
tagacaccta tggccggata actaaggatg tggtgctcca atggggagaa atgcccacgt 3300
ctgtggccta cattcattcc gatcagataa tgggctgggg cgagaaagct attgagatcc 3360
ggtcagtgga aacaggacat ttggatggag tatttatgca taagcgagct caaaggttaa 3420
agtttctatg tgaaagaaat gataaggtat tttttgcatc cgtgcgatct ggaggaagta 3480
gccaagtgtt tttcatgacc ctcaacagaa attccatgat gaactggtaa cagaagagca 3540
cttggcactt atcttcatgg cgttatttct aatttaaaag aacataactc atgtggactt 3600
atgccagtct agaggcagaa tcagaaggct tggttgaaca tatcgctttc ccttttttcct 3660
ctccctccgc ccctcccagt acagtccatc tttcaatgtt gcagcctggt tgagaaggag 3720
agaaaaaggt ggcaggaatt tccaggagat ccccaagaat gctgccttgt ctgtggacaa 3780
agatggacca tgtgcccttc ggaattaggg atagaaacaa atattgtgtg ctcttaacga 3840
ttaagctgtg ttatggtggg ttttcaggtt tttacctttt ttctttaccc ctctactctg 3900
caagaatggg gaaagaatgc atactgcgca aatgagtctt ttaaattctg tctg 3954

<210> 27
<211> 3357
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7517148CB1

<400> 27
agtgtgctgg aattcgcgcc ttttacgcca aagctcgcca acatggcgga cctggaggct 60
gtgctggccg atgtcagtta cctgatggcc atggagaaga gcaaggcgac cccggccgcc 120
cgcgccagca agaggatcgt cctgccggag cccagtatcc ggagtgtgat gcagaagtac 180
cttgcagaga gaaatgaaat aacccttgac aagattttca atcagaaaat tggtttcttg 240
ctatttaaag atttttgttt gaatgaaatt aatgaagctg tacctcaggt gaagttttat 300
gaagagataa aggaatatga aaaacttgat aatgaggaag accgcctttg cagaagtcga 360
caaatttatg atgcctacat catgaaggaa cttctttcct gttcacatcc tttctcaaag 420
caagctgtag aacacgtaca aagtcattta tccaagaaac aagtgacatc aactcttttt 480
cagccataca tagaagaaat ttgtgaaagc cttcgaggtg acatttttca aaaatttatg 540
gaaagtgaca agttcactag attttgtcag tggaaaaacg ttgaattaaa tatccatttg 600

PF-1506 P

accatgaatg agttcagtgt gcataggatt attggacgag gaggattcgg ggaagtttat 660
ggttgcagga aagcagacac tggaaaaatg tatgcaatga aatgcttaga taagaagagg 720
atcaaaatga aacaaggaga aacattagcc ttaaatgaaa gaatcatgtt gtctcttgtc 780
agcacaggag actgtccttt cattgtatgt atgacctatg ccttccatac cccagataaa 840
ctctgcttca tcctggatct gatgaacggg ggcgatttgc actaccacct ttcacaacac 900
ggtgtgttct ctgagaagga gatgcggttt tatgccactg aaatcattct gggtctggaa 960
cacatgcaca atcggtttgt tgtctacaga gatttgaagc cagcaaatat tctcttggat 1020
gaacatggac acgcaagaat atcagatctt ggtcttgcct gcgatttttc caaaaagaag 1080
cctcatgcga gtgttggcac ccatgggtac atggctcccg aggtgctgca gaaggggacg 1140
gcctatgaca gcagtgccga ctggttctcc ctgggctgca tgcttttcaa acttctgaga 1200
ggtcacagcc ctttcagaca acataaaacc aaagacaagc atgaaattga ccgaatgaca 1260
ctcaccgtga atgtggaact tccagacacc ttctctcctg aactgaagtc ccttttggag 1320
ggcttgcttc agcgagacgt tagcaagcgg ctgggctgtc acggaggcgg ctcacaggaa 1380
gtaaaagagc acagcttttt caaaggtgtt gactggcagc atgtctactt acaaaagtac 1440
ccaccaccct tgattcctcc ccggggagaa gtcaatgctg ctgatgcctt tgatattggc 1500
tcatttgatg aagaggatac caaagggatt aagctacttg attgcgacca agaactctac 1560
aagaacttcc ctttggtcat ctctgaacgc tggcagcaag aagtaacgga aacagtttat 1620
gaagcagtaa atgcagacac agataaaatc gaggccagga agagagctaa aaataagcaa 1680
cttggccacg aagaagatta cgctctgggg aaggactgta ttatgcacgg gtacatgctg 1740
aaactgggaa acccatttct gactcagtgg cagcgtcgct atttttacct ctttccaaat 1800
agacttgaat ggagaggaga gggagagtcc cggagtgatc cagagtttgt gcagtggaag 1860
aaagagttga acgaaacctt caaggaggcc cggcggctat tgcgtcgtgc cccgaagttc 1920
ctcaacaaac ctcggtcagg tactgtggag ctcccaaagc catccctctg tcacagaaac 1980
agcaacggcc tctagcaccc agaaacaggg agggtcctcg aggaggacac accagggtct 2040
cagcctttg gggtgaacga ggatgaggca tctgatctat tcgctaccgg gactcctcca 2100
ggctcccgag aggagtcggg acccttcggc ttggggtcag ctcagctccc tgccttgtca 2160
catttgtctg cattagaaac tactgaagaa ataaaagttc tttttctttg ctacacactt 2220
tggtacctat gaacctagaa cttgaagtga ctcctactta tcacgtaaat ttttatgtct 2280
gatatcaaac acatcttaga ctccccagaa tggaatttaa agatgttcag tgttgggtaa 2340
cagattgccc taagcattgc cacatattct gtctagtcgc tgctgatttt ctatgtcttt 2400
gctccatact gcaggggat gggagagcca cagtgtgttt cttttgtgca cttcgcaact 2460
gacttcttgt cctggggtta aaagttgaag atattttctg atgatattaa aagttgaaga 2520
tatttctgca cttgggccct cctctgggag ccgcacccac atgactgccc tgcctctgac 2580
cagtctgttc cggggccccc tcagccaggt gggaatgacg gacacgtact atccaagtgt 2640
atgggattaa ctaatcattg aaggcattca tccgtccatc attggaaaga tttacagtga 2700
ttctgaagga caggccgtgg agttttaggt ttcaggggca agagcagttt tcaaaagtct 2760
ttgagtccag tgtgcacgag tcgacaagca gtacctggca tgcaggagca ctcatgggtg 2820
agtccgtctc aggtctcgac aattagcagt tgtgtgacag tcattctggt tccttctgcc 2880
tgaccctggg agacatatca gtaatggatg tacaaaagca ggtctgtttt atgtcttagt 2940
ataatttcag atgaattgta ttgaaaaaat gctgaggaat gaatgtgtca aaatgggtta 3000
actgtgtata ttgactttca tgtcgtcatg catctgtcat gaatgaatga tactttgcac 3060
tgggctgtac gacagtgagg accttagggc atgaagcctt tttcctggtc ccagcagcat 3120
ctgccctgtg aagtttgttt tctcccactg cctccaggcc ccactgatac ccccaaatag 3180
atgctgggtt atgagaacca gcgaaatccc ccatgtcatc agtcttaaaa aaaaaatttt 3240
acaaatccac gtatttgtcc cattcttgga gtagttttag tgtatgtctt tacattaact 3300
actaacagta taaataactt gacatcgtaa ttgtctgcat cctgtccttg atatttt 3357

<210> 28
<211> 2036
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7517238CB1

<400> 28
aagattctta ggaacgccgt acctgccgcg tctctcagga cagcaggccc ctgtccttct 60
gtcgggcgcc gctcagccgt gccctccgcc cctcaggttc tttttctaat tccaaataaa 120
cttgcaagag gactatgaaa gattatgatg aacttctcaa atattatgaa ttacatgaaa 180
ctattgggac aggtggcttt gcaaaggtca aacttgcctg ccatatcctt actggagaga 240
tggtagctat aaaaatcatg gataaaaaca cactagggag tgatttgccc cggatcaaaa 300
cggagattga ggccttgaag aacctgagac atcagcatat atgtcaactc taccatgtgc 360
tagagacagc caacaaaata ttcatggttc ttgaggaaaa tttgctgttt gatgaatatc 420
ataaattaaa gctgattgac tttggtctct gtgcaaaacc caagggtaac aaggattacc 480
atctacagac atgctgtggg agtctggctt atgcagcacc tgagttaata caaggcaaat 540

PF-1506 P

```
catatcttgg atcagaggca gatgtttgga gcatgggcat actgttatat gttcttatgt 600
gtggatttct accatttgat gatgataatg taatggcttt atacaagaag attatgagag 660
gaaaatatga tgttcccaag tggctctctc ccagtagcat tctgcttctt caacaaatgc 720
tgcaggtgga cccaaagaaa cggatttcta tgaaaaatct attgaaccat ccctggatca 780
tgcaagatta caactatcct gttgagtggc aaagcaagaa tccttttatt cacctcgatg 840
atgattgcgt aacagaactt tctgtacatc acagaaacaa caggcaaaca atggaggatt 900
taatttcact gtggcagtat gatcacctca cggctaccta tcttctgctt ctagccaaga 960
aggctcgggg aaaaccagtt cgtttaaggc tttcttcttt ctcctgtgga caagccagtg 1020
ctaccccatt cacagacatc aagtcaaata attggagtct ggaagatgtg accgcaagtg 1080
ataaaaatta tgtggcggga ttaatagact atgattggtg tgaagatgat ttatcaacag 1140
gtgctgctac tccccgaaca tcacagttta ccaagtactg gacagaatca aatggggcgg 1200
aatctaaatc attaactcca gccttatgca gaacacctgc aaataaatta aagaacaaag 1260
aaaatgtata tactcctaag tctgctgtaa agaatgaaga gtactttatg tttcctgagc 1320
caaagactcc agttaataag aaccagcata agagagaaat actcactacg ccaaatcgtt 1380
acactacacc ctcaaaagct agaaaccagt gcctgaaaga aactccaatt aaaataccag 1440
taaattcaac aggaacagac aagttaatga caggtgtcat tagccctgag aggcggtgcc 1500
gctcagtgga attggatctc aaccaagcac atatggagga gactccaaaa agaaagggag 1560
ccaaagtgtt tgggagcctt gaaaggggct tggataaggt tatcactgtg ctcaccagga 1620
gcaaaaggaa gggttctgcc agagacgggc ccagaagact aaagcttcac tataatgtga 1680
ctacaactag attagtgaat ccagatcaac tgttgaatga aataatgtct attcttccaa 1740
agaagcatgt tgactttgta caaaagggtt atacactgaa gtgtcaaaca cagtcagatt 1800
ttgggaaagt gacaatgcaa tttgaattag aagtgtgcca gcttcaaaaa cccgatgtgg 1860
tgggtatcag gaggcagcgg cttaagggcg atgcctgggt ttacaaaaga ttagtggaag 1920
acatcctatc tagctgcaag gtataattga tggattcttc catcctgccg gatgagtgtg 1980
ggtgtgatac agcctacata aagactaaga gcgaattcgc agcacactga cgccgg     2036
```

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7517831	1	7517831CD1	15	7517831CB1
7520272	2	7520272CD1	16	7520272CB1
7521279	3	7521279CD1	17	7521279CB1
7523965	4	7523965CD1	18	7523965CB1
7524016	5	7524016CD1	19	7524016CB1
7524680	6	7524680CD1	20	7524680CB1
7524757	7	7524757CD1	21	7524757CB1
7516229	8	7516229CD1	22	7516229CB1
7516525	9	7516525CD1	23	7516525CB1
7516533	10	7516533CD1	24	7516533CB1
7516613	11	7516613CD1	25	7516613CB1
7517068	12	7517068CD1	26	7517068CB1
7517148	13	7517148CD1	27	7517148CB1
7517238	14	7517238CD1	28	7517238CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7517831CD1	g775208	4.3E-21	[Homo sapiens] p56lck
				Vogel, L. B. et al., p70 phosphorylation and binding to p56lck is an early event in interleukin-2-induced onset of cell cycle progression in T-lymphocytes, J. Biol. Chem. 270, 2506-2511 (1995)
		342146|LCK	7.2E-20	[Homo sapiens][Protein kinase;Transferase] Lymphocyte-specific protein tyrosine kinase, tyrosine kinase that is involved in T cell receptor signaling through Ras and MAPK pathways, activator of CASP8 in radiation-induced apoptosis; gene defect correlates with immunodeficiency plus CD4 lymphopenia
				Su, S. B. et al., Inhibition of tyrosine kinase activation blocks the down-regulation of CXC chemokine receptor 4 by HIV-1 gp120 in CD4+ T cells, J Immunol 162, 7128-32 (1999).
		780711|Lck	8.0E-17	[Mus musculus][Protein kinase;Transferase] Lymphocyte-specific protein tyrosine kinase, tyrosine kinase that is involved in T cell receptor signaling through Ras and MAPK pathways, regulates T cell development and apoptosis; human gene defect correlates with immunodeficiency plus CD4 lymphopenia
				Legname, G. et al., Inducible expression of a p56Lck transgene reveals a central role for Lck in the differentiation of CD4 SP thymocytes, Immunity 12, 537-46 (2000).
2	7520272CD1	g439226	4.0E-152	[Homo sapiens] fructose-1,6-bisphosphatase
				Kikawa, Y. et al., cDNA sequences encoding human fructose 1,6-bisphosphatase from monocytes, liver and kidney: Application of monocytes to molecular analysis of human fructose 1,6-bisphosphatase deficiency, Cell. Mol. Biol. Res. 199, 687-693 (1994)
		753731|FBP1	3.0E-153	[Homo sapiens][Other phosphatase;Hydrolase] Fructose-1,6-bisphosphatase 1 (liver), hydrolyzes fructose-1,6-bisphosphate to fructose-6-phosphate and inorganic phosphate, regulatory step in gluconeogenesis; deficiency is associated with metabolic acidosis and fasting hypoglycemia

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				el-Maghrabi, M. R. et al., Isolation of a human liver fructose-1,6-bisphosphatase cDNA and expression of the protein in Escherichia coli. Role of ASP-118 and ASP- 121 in catalysis, J Biol Chem 268, 9466-72 (1993).
3	7521279CD1	g1905761	4.1E-233	[Homo sapiens] 6-phosphofructo-2-kinase/fructose-2, 6-bisphosphatase
				Sakai, A. et al., Cloning of cDNA encoding for a novel isozyme of fructose 6-phosphate, 2-kinase/fructose 2,6-bisphosphatase from human placenta, J. Biochem. 119, 506-511 (1996)
		341042|PFKFB4	3.0E-234	[Homo sapiens][Transferase;Other kinase;Other phosphatase;Hydrolase] 6-phosphofructo-2-kinase, fructose-2,6-biphosphatase, testis form, synthesizes and degrades fructose-2,6-bisphosphate and may be involved in the regulation of glycolysis
				Manzano, A. et al., Cloning, expression and chromosomal localization of a human testis 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase gene, Gene 229, 83-9 (1999).
		609815|Pfkfb4	1.1E-227	[Rattus norvegicus][Transferase;Other kinase;Other phosphatase;Hydrolase] 6-phosphofructo-2-kinase, fructose-2,6-biphosphatase, testis form, synthesizes and degrades fructose-2,6-bisphosphate and may be involved in the regulation of glycolysis
				Li, L. et al., Expression of chicken liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase in Escherichia coli, Biochem Biophys Res Commun 209, 883-93 (1995).
4	7523965CD1	g2661752	1.1E-120	[Homo sapiens] phosphoenolpyruvate carboxykinase (GTP)
				Modaressi, S. et al., Human mitochondrial phosphoenolpyruvate carboxykinase 2 gene. Structure, chromosomal localization and tissue-specific expression, Biochem. J. 333 (Pt 2), 359-366 (1998)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		341026|PCK2	5.7E-121	[Homo sapiens][Lyase;Other kinase][Cytoplasmic;Mitochondrial] Phosphoenolpyruvate carboxykinase 2, catalyzes the formation of phosphoenolpyruvate by decarboxylation of oxaloacetate, rate-limiting step of gluconeogenesis
				Modaressi, S. et al., Molecular cloning, sequencing and expression of the cDNA of the mitochondrial form of phosphoenolpyruvate carboxykinase from human liver, Biochem J 315, 807-14 (1996).
		586739|Pck1	2.6E-68	[Mus musculus][Lyase;Other kinase][Cytoplasmic] Cytosolic phosphoenolpyruvate carboxykinase, catalyzes the formation of phosphoenolpyruvate by decarboxylation of oxaloacetate
				She, P. et al., Phosphoenolpyruvate carboxykinase is necessary for the integration of hepatic energy metabolism, Mol Cell Biol 20, 6508-17 (2000).
5	7524016CD1	g35503	1.7E-94	[Homo sapiens] 6-phosphofructo-2-kinase/fructose-2,6- bisphosphatase (AA 1-471)
				Lange, A. J. et al., Sequence of human liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase, Nucleic Acids Res. 18, 3652 (1990)
		336898|PFKFB1	1.2E-95	[Homo sapiens][Protein phosphatase;Transferase;Other phosphatase;Other kinase;Hydrolase] 6-phosphofructo-2-kinase, fructose-2,6-biphosphatase 1, liver and muscle form, enzyme involved in regulating glycolysis, catalyzes the synthesis and degradation of fructose-2,6-bisphosphate
				Lange, A. J. et al., Expression of human liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase in Escherichia coli. Role of N-2 proline in degradation of the protein, J Biol Chem 268, 8078-84 (1993).
		430618|Pfkfb1	9.3E-89	[Rattus norvegicus][Protein phosphatase;Transferase;Other phosphatase;Other kinase;Hydrolase] 6-phosphofructo-2-kinase, fructose-2,6-biphosphatase 1, liver and muscle form, enzyme involved in regulating glycolysis, catalyzes the synthesis and degradation of fructose-2,6-bisphosphate

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Kurland, I. J. et al., Rat liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. Properties of phospho- and dephospho- forms and of two mutants in which Ser32 has been changed by site-directed mutagenesis, J Biol Chem 267, 4416-23 (1992).
6	7524680CD1	g35503	1.3E-215	[Homo sapiens] 6-phosphofructo-2-kinase/fructose-2,6- bisphosphatase (AA 1-471)
				Lange, A. J. et al., Sequence of human liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase, Nucleic Acids Res. 18, 3652 (1990)
		336898|PFKFB1	9.1E-217	[Homo sapiens][Protein phosphatase;Transferase;Other phosphatase;Other kinase;Hydrolase] 6-phosphofructo-2-kinase, fructose-2,6-biphosphatase 1, liver and muscle form, enzyme involved in regulating glycolysis, catalyzes the synthesis and degradation of fructose-2,6-bisphosphate
				Lange, A. J. et al., Expression of human liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase in Escherichia coli. Role of N-2 proline in degradation of the protein, J Biol Chem 268, 8078-84 (1993).
		430618|Pfkfb1	1.2E-207	[Rattus norvegicus][Protein phosphatase;Transferase;Other phosphatase;Other kinase;Hydrolase] 6-phosphofructo-2-kinase, fructose-2,6-biphosphatase 1, liver and muscle form, enzyme involved in regulating glycolysis, catalyzes the synthesis and degradation of fructose-2,6-bisphosphate
				Kurland, I. J. et al., Rat liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. Properties of phospho- and dephospho- forms and of two mutants in which Ser32 has been changed by site-directed mutagenesis, J Biol Chem 267, 4416-23 (1992).
7	7524757CD1	g35503	3.7E-223	[Homo sapiens] 6-phosphofructo-2-kinase/fructose-2,6- bisphosphatase (AA 1-471)
				Lange, A. J. et al., Sequence of human liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase, Nucleic Acids Res. 18, 3652 (1990)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		336898|PFKFB1	2.7E-224	[Homo sapiens][Protein phosphatase;Transferase;Other phosphatase;Other kinase;Hydrolase] 6-phosphofructo-2-kinase, fructose-2,6-biphosphatase 1, liver and muscle form, enzyme involved in regulating glycolysis, catalyzes the synthesis and degradation of fructose-2,6-bisphosphate
				Lange, A. J. et al., Expression of human liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase in Escherichia coli. Role of N-2 proline in degradation of the protein, J Biol Chem 268, 8078-84 (1993).
		430618|Pfkfb1	3.7E-216	[Rattus norvegicus][Protein phosphatase;Transferase;Other phosphatase;Other kinase;Hydrolase] 6-phosphofructo-2-kinase, fructose-2,6-biphosphatase 1, liver and muscle form, enzyme involved in regulating glycolysis, catalyzes the synthesis and degradation of fructose-2,6-bisphosphate
				Kurland, I. J. et al., Rat liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. Properties of phospho- and dephospho- forms and of two mutants in which Ser32 has been changed by site-directed mutagenesis, J Biol Chem 267, 4416-23 (1992).
8	7516229CD1	g6760472	3.0E-190	[Homo sapiens] type II phosphatidylinositol-4-phosphate 5-kinase 53K isoform
				Boronenkov, I. V. et al., The sequence of phosphatidylinositol-4-phosphate 5-kinase defines a novel family of lipid kinases, J. Biol. Chem. 270, 2881-2884 (1995)
		568490|PIP5K2A	2.1E-191	[Homo sapiens][Transferase;Other kinase] Phosphatidylinositol-4-phosphate 5-kinase type II, alpha, a member of a family of kinases responsible for the synthesis of PtdIns(4,5)P2
				Boronenkov, I. V. et al., The sequence of phosphatidylinositol-4-phosphate 5-kinase defines a novel family of lipid kinases, J Biol Chem 270, 2881-4 (1995).
		757680|Pip5k2a	5.0E-188	[Rattus norvegicus] Phosphatidylinositol-4-phosphate 5-kinase type II alpha
				Itoh, T. et al., Autophosphorylation of type I phosphatidylinositol phosphate kinase regulates its lipid kinase activity, J Biol Chem 275, 19389-94 (2000).
9	7516525CD1	g23499314	6.7E-270	[Homo sapiens] (AF425232) CaMKK alpha protein

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		716531|DKFZp761M0423	4.6E-271	[Homo sapiens] Protein with strong similarity to calcium-calmodulin-dependent protein kinase kinase 1 alpha (rat Camkk1), which phosphorylates and activates Ca(2+)-calmodulin (CaM)-dependent kinase I and IV but not CaM kinase II, contains a protein kinase domain
		711580|Camkk1	2.4E-254	[Rattus norvegicus][Protein kinase;Transferase] Calcium-calmodulin-dependent protein kinase kinase 1 alpha, phosphorylates and activates Ca(2+)-calmodulin (CaM)-dependent kinase I and IV but not CaM kinase II, involved in Ca(2+)-calmodulin signaling
				Okuno, S. et al., Regulation of Ca(2+)/Calmodulin-Dependent Protein Kinase Kinase alpha by cAMP-Dependent Protein Kinase: I. Biochemical Analysis, J Biochem (Tokyo) 130, 503-13. (2001).
10	7516533CD1	g189508	3.6E-240	[Homo sapiens] p70 ribosomal S6 kinase alpha-I
				Grove, J. R. et al., Cloning and expression of two human p70 S6 kinase polypeptides differing only at their amino termini, Mol. Cell. Biol. 11, 5541-5550 (1991)
		337822|RPS6KB1	2.5E-241	[Homo sapiens][Protein kinase;Transferase] Ribosomal protein S6 kinase, 70kD, a member of the ribosomal protein S6 kinase (RSK) family of protein kinases, insulin and mitogen activated, and plays roles in cell cycle progression and control of cell proliferation
				Brenneisen, P. et al., Activation of p70 ribosomal protein S6 kinase is an essential step in the DNA damage-dependent signaling pathway responsible for the ultraviolet B-mediated increase in interstitial collagenase (MMP-1) and stromelysin-1 (MMP-3) protein levels in human dermal fibroblasts, J Biol Chem 275, 4336-44. (2000).
		711952|Rps6kb1	5.3E-239	[Rattus norvegicus][Protein kinase;Transferase] Ribosomal protein S6 kinase, 70kD, a member of the ribosomal protein S6 kinase (RSK) family of protein kinases, insulin and mitogen activated, and plays roles in cell cycle progression and control of cell proliferation

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Grove, J. R. et al., Cloning and expression of two human p70 S6 kinase polypeptides differing only at their amino termini, Mol Cell Biol 11, 5541-50 (1991).
11	7516613CD1	g1872546	0.0	[Mus musculus] NIK
				Su, Y. C. et al., NIK is a new Ste20-related kinase that binds NCK and MEKK1 and activates the SAPK/JNK cascade via a conserved regulatory domain, EMBO J. 16, 1279-1290 (1997)
		582239|Map4k4	0.0	[Mus musculus][Protein kinase;Transferase;Receptor (signalling)] Mitogen-activated protein kinase kinase kinase kinase 4, a serine-threonine kinase, interacts with Nck, interacts with MEKK1 (Map3k1) and activates the c-Jun N-terminal kinase (Mapk8) signaling pathway; mutants fail to develop somites or a hindgut
				Becker, E. et al., Nck-interacting Ste20 kinase couples Eph receptors to c-Jun N-terminal kinase and integrin activation, Mol Cell Biol 20, 1537-45. (2000).
		340694|MAP4K4	0.0	[Homo sapiens][Protein kinase;Transferase] Mitogen-activated protein kinase kinase kinase kinase 4, a serine-threonine kinase, activates the c-Jun N-terminal kinase (MAPK8) signaling pathway, does not activate the ERK or p38 (CSBP1) kinase pathways, may be involved in TNF-alpha (TNF) signaling
				Yao, Z. et al., A novel human STE20-related protein kinase, HGK, that specifically activates the c-Jun N-terminal kinase signaling pathway, J Biol Chem 274, 2118-25 (1999).
12	7517068CD1	g6110362	0.0	[Homo sapiens] Traf2 and NCK interacting kinase, splice variant 7
				Fu, C. A. et al., TNIK, a novel member of the germinal center kinase family that activates the c-Jun N-terminal kinase pathway and regulates the cytoskeleton, J. Biol. Chem. 274, 30729-30737 (1999)
		340694|MAP4K4	0.0	[Homo sapiens][Protein kinase;Transferase] Mitogen-activated protein kinase kinase kinase kinase 4, a serine-threonine kinase, activates the c-Jun N-terminal kinase (MAPK8) signaling pathway, does not activate the ERK or p38 (CSBP1) kinase pathways, may be involved in TNF-alpha (TNF) signaling

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Yao, Z. et al., A novel human STE20-related protein kinase, HGK, that specifically activates the c-Jun N-terminal kinase signaling pathway, J Biol Chem 274, 2118-25 (1999).
		582239|Map4k4	0.0	[Mus musculus][Protein kinase;Transferase;Receptor (signalling)] Mitogen-activated protein kinase kinase kinase kinase 4, a serine-threonine kinase, interacts with Nck, interacts with MEKK1 (Map3k1) and activates the c-Jun N-terminal kinase (Mapk8) signaling pathway; mutants fail to develop somites or a hindgut
				Su, Y. C. et al., NIK is a new Ste20-related kinase that binds NCK and MEKK1 and activates the SAPK/JNK cascade via a conserved regulatory domain, Embo Journal 16, 1279-90 (1997).
13	7517148CD1	g312395	0.0	[Homo sapiens] beta-adrenergic kinase 2
				Parruti, G. et al., Molecular cloning, functional expression and mRNA analysis of human beta-adrenergic receptor kinase 2, Biochem. Biophys. Res. Commun. 190, 475-481 (1993)
		341946|ADRBK2	0.0	[Homo sapiens][Protein kinase;Transferase][Cytoplasmic;Plasma membrane] G-protein coupled receptor kinase 3, member of a family of protein kinases that specifically phosphorylate activated G protein coupled receptors, resulting in receptor desensitization, may represent a genetic marker for mood disorders
				Parruti, G. et al., Molecular cloning, functional expression and mRNA analysis of human beta-adrenergic receptor kinase 2, Biochem Biophys Res Commun 190, 475-81 (1993).
		589791|Adrbk2	0.0	[Rattus norvegicus][Protein kinase;Transferase][Axon;Dense bodies] G-protein coupled receptor kinase 3, member of a family of protein kinases that specifically phosphorylate activated G protein coupled receptors, resulting in receptor desensitization, may regulate nociception, sperm chemotaxis and olfaction

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Kovoor, A. et al., Agonist induced homologous desensitization of mu-opioid receptors mediated by G protein-coupled receptor kinases is dependent on agonist efficacy, Mol Pharmacol 54, 704-11 (1998).
14	7517238CD1	g15559349	0.0	[Homo sapiens] Similar to likely ortholog of maternal embryonic leucine zipper kinase
		570006|MELK	0.0	[Homo sapiens][Protein kinase;Transferase] Protein containing two C-terminal kinase associated domain 1 and two protein kinase domains, has low similarity to microtubule-MAP-affinity regulating kinase (rat LOC60328), which is a serine-threonine kinase that influences microtubule stability
				Seong, H. A. et al., Phosphorylation of a novel zinc-finger-like protein, ZPR9, by murine protein serine/threonine kinase 38 (MPK38), Biochem J 361, 597-604. (2002).
		585291|Melk	6.8E-270	[Mus musculus][Protein kinase;Transferase] Protein containing a protein kinase domain and a C-terminal kinase associated domain 1, has low similarity to rat LOC60328, which is a serine-threonine kinase that participates in microtubule stability and the control of cell polarity
				Seong, H. A. et al., Phosphorylation of a novel zinc-finger-like protein, ZPR9, by murine protein serine/threonine kinase 38 (MPK38), Biochem J 361, 597-604. (2002).

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7517831CD1	83	signal_cleavage: M1-T58	SPSCAN
			KINASE TYROSINE-PROTEIN PROTO-ONCOGENE DOMAIN TRANSFERASE ATP-BINDING MYRISTATE PHOSPHORYLATION SH2 SH3 PD012180: G2-E43	BLAST_PRODOM
			Potential Phosphorylation Sites: S7	MOTIFS
			Potential Glycosylation Sites: N40, N67	MOTIFS
2	7520272CD1	292	signal_cleavage: M1-A44	SPSCAN
			Fructose-1-6-bisphosphatase: N12-H289	HMMER_PFAM
			Inositol phosphatase/fructose-1,6-bisphosphatase IPB000146: G59-D100, G112-T135, Q155-P189, R198-P220, G228-G253	BLIMPS_BLOCKS
			Fructose-1-6-bisphosphatase active site: H208-E255	PROFILESCAN
			Fructose-1,6-bisphosphatase signature PR00115: D119-Y140, P156-L176, G181-G196, A197-P220, G228-G248, V257-V279	BLIMPS_PRINTS
			Inositol phosphatase/fructose-1,6-bisphosphatase family signature PR00377: V115-N126, L211-A221, Y234-G248	BLIMPS_PRINTS
			HYDROLASE CARBOHYDRATE METABOLISM FRUCTOSE-16-BISPHOSPHATASE FBPASE 1-PHOSPHOHYDROLASE D-FRUCTOSE-1,6-BISPHOSPHATE CYCLE CHLOROPLAST CALVIN PD001491:G68-P189 D188-V279	BLAST_PRODOM
			FRUCTOSE1 6-BISPHOSPHATASE 1-PHOSPHOHYDROLASE FBPASE HYDROLASE CARBOHYDRATE METABOLISM D-FRUCTOSE-1, 6-BISPHOSPHATE MULTI-GENE PD017713: T13-V66	BLAST_PRODOM
			FRUCTOSE-1-6-BISPHOSPHATASE DM00535|P09467|10-331:V11-P189 P189-E287	BLAST_DOMO
			FRUCTOSE-1-6-BISPHOSPHATASE DM00535|A37295|60-331:A61-P189 P189-E287	BLAST_DOMO
			FRUCTOSE-1-6-BISPHOSPHATASE DM00535|S46245|11-332:T13-P189 D190-G274	BLAST_DOMO
			FRUCTOSE-1-6-BISPHOSPHATASE DM00535|P46267|12-333:T13-P189 D190-Y286	BLAST_DOMO
			Potential Phosphorylation Sites: S97, S125, S144, S149, S275, T145, T252	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Glycosylation Sites: N65	MOTIFS
			Fructose-1-6-bisphosphatase active site: G228-A240	MOTIFS
3	7521279CD1	434	signal_cleavage: M1-G51	SPSCAN
			6-phosphofructo-2-kinase: Q30-P249	HMMER_PFAM
			Phosphoglycerate mutase family: R250-I400	HMMER_PFAM
			Phosphoglycerate mutase family IPB001345: I252-S284, V299-A311, G315-R347	BLIMPS_BLOCKS
			Phosphoglycerate mutase family phosphohistidine signature: I234-K283	PROFILESCAN
			6-phosphofructo-2-kinase family signature PR00991: V125-A139, K151-I165, P177-F191, V230-S251, I252-L274	BLIMPS_PRINTS
			MUTASE PROTEOME COMPLETE PHOSPHOGLYCERATE PGAM ISOMERASE GLYCOLYSIS BPG-DEPENDENT FRUCTOSE-2,6-BISPHOSPHATASE PHOSPHOGLYCEROMUTASE PD000730:Y253-D328 S330-L388	BLAST_PRODOM
			KINASE FRUCTOSE-2,6-BISPHOSPHATASE INCLUDES: ISOZYME 6PF-2-K/FRU- 6-PHOSPHOFRUCTO-2-KINASE/FRUCTOSE-2,6-BIPHOSPHATASE TRANSFERASE 2,6-P2ASE MULTI-FUNCTIONAL PD002665: T36-I252	BLAST_PRODOM
			6-BISPHOSPHATASE TRANSFERASE 6PF2K/FRU2 6-P2ASE INCLUDES: KINASE FRUCTOSE2 MULTI-FUNCTIONAL ENZYME ISOZYME PD009472: T389-H433	BLAST_PRODOM
			6-PF2K/FRU2,6-P2ASE TESTIS ISOZYME INCLUDES: 6-PHOSPHOFRUCTO 2-KINASE EC 2.7.1.105 FRUCTOSE-2,6-BISPHOSPHATASE 3.1.3.46 MULTI-FUNCTIONAL ENZYME TRANSFERASE KINASE HYDROLASE ATP-BINDING PD114268: M1-M35	BLAST_PRODOM
			6-PHOSPHOFRUCTO-2-KINASE / FRUCTOSE-2,6-BISPHOSPHATE 2-PHOSPHATASE DM01656|JC1470|184-441:E186-R337 Y331-V407	BLAST_DOMO
			6-PHOSPHOFRUCTO-2-KINASE / FRUCTOSE-2,6-BISPHOSPHATE 2-PHOSPHATASE DM01656|JC2037|185-444:N181-L352 S330-V407	BLAST_DOMO
			6-PHOSPHOFRUCTO-2-KINASE / FRUCTOSE-2,6-BISPHOSPHATE 2-PHOSPHATASE DM01656|P07953|184-442: G46-V82 D183-A329 S330-V407	BLAST_DOMO
			6-PHOSPHOFRUCTO-2-KINASE / FRUCTOSE-2,6-BISPHOSPHATE 2-PHOSPHATASE DM01656|P25114|183-441:D183-E349 S330-V407	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S3, S56, S204, S275, S330, T60, T85, T133, T140, T248, T409, Y377	MOTIFS
			Potential Glycosylation Sites: N132	MOTIFS
			ATP/GTP-binding site motif A (P-loop): G46-T53	MOTIFS
			Phosphoglycerate mutase family phosphohistidine signature: L254-N263	MOTIFS
4	7523965CD1	240	Phosphoenolpyruvate carboxykinase (GTP) IPB000364: K88-P121, F148-L178, T179-L202, D204-P217	BLIMPS_BLOCKS
			PHOSPHOENOLPYRUVATE CARBOXYKINASE LYASE GTP-BINDING CARBOXYLASE DECARBOXYLASE GLUCONEOGENESIS PD004738: D46-E232	BLAST_PRODOM
			PHOSPHOENOLPYRUVATE CARBOXYKINASE, MITOCHONDRIAL PRECURSOR GTP EC 4.1.1.32 CARBOXYLASE PEPCKM GLUCONEOGENESIS LYASE DECARBOXYLASE GTP-BINDING MITOCHONDRION TRANSIT PEPTIDE MANGANES PD144568: M1-R45	BLAST_PRODOM
			PHOSPHOENOLPYRUVATE CARBOXYKINASE (GTP) DM01781|P05153|15-621: V32-P240 DM01781|P20007|40-646: G35-E232 DM01781|P21642|33-639: L33-P240 DM01781|Q05893|30-640: V32-E232	BLAST_DOMO
			Potential Phosphorylation Sites: S23, S51, S115, S136, S187, T29, T66, T75, T219	MOTIFS
5	7524016CD1	199	signal_cleavage: M1-T33	SPSCAN
			6-phosphofructo-2-kinase: R7-W199	HMMER_PFAM
			6-phosphofructo-2-kinase family signature PR00991: V104-S118, K130-I144, P156-F170	BLIMPS_PRINTS
			KINASE FRUCTOSE-2,6-BISPHOSPHATASE INCLUDES: ISOZYME 6PF-2-K/FRU- 6-PHOSPHOFRUCTO-2-KINASE/FRUCTOSE-2,6-BIPHOSPHATASE TRANSFERASE 2,6-P2ASE MULTI-FUNCTIONAL PD002665: W45-W194, A10-A124	BLAST_PRODOM
			6-PHOSPHOFRUCTO-2-KINASE / FRUCTOSE-2,6-BISPHOSPHATE 2-PHOSPHATASE DM01457|JC1470|28-182: A10-C161 DM01457|P07953|29-182: A10-C161 DM01457|P25114|27-181: T16-D160 DM01457|P26285|26-180: T16-Y158	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S36, S64, S98, T5, T112	MOTIFS
			Potential Glycosylation Sites: N111	MOTIFS
			ATP/GTP-binding site motif A (P-loop): G26-T33	MOTIFS
6	7524680CD1	406	6-phosphofructo-2-kinase: M1-P186	HMMER_PFAM
			Phosphoglycerate mutase family: R187-I372	HMMER_PFAM
			Phosphoglycerate mutase family IPB001345: I189-A221, V236-A248, G252-E284, E301-E346	BLIMPS_BLOCKS
			Phosphoglycerate mutase family phosphohistidine signature: I171-Y220	PROFILESCAN
			6-phosphofructo-2-kinase family signature PR00991: V62-S76, K88-I102, P114-F128, V167-S188, I189-L211, A266-P282	BLIMPS_PRINTS
			MUTASE PROTEOME COMPLETE PHOSPHOGLYCERATE PGAM ISOMERASE GLYCOLYSIS BPG-DEPENDENT FRUCTOSE-2,6-BISPHOSPHATASE PHOSPHOGLYCEROMUTASE PD000730:Y190-Y303 P298-L360	BLAST_PRODOM
			KINASE FRUCTOSE-2,6-BISPHOSPHATASE INCLUDES: ISOZYME 6PF-2-K/FRU- 6-PHOSPHOFRUCTO-2-KINASE/FRUCTOSE-2,6-BIPHOSPHATASE TRANSFERASE 2,6-P2ASE MULTI-FUNCTIONAL PD002665: K9-I189	BLAST_PRODOM
			6-BISPHOSPHATASE TRANSFERASE 6PF2K/FRU2 6-P2ASE INCLUDES: KINASE FRUCTOSE2 MULTI-FUNCTIONAL ENZYME ISOZYME PD009472: T361-Y406	BLAST_PRODOM
			FRUCTOSE-2 SIMILAR PD114271: S232-V376	BLAST_PRODOM
			6-PHOSPHOFRUCTO-2-KINASE / FRUCTOSE-2,6-BISPHOSPHATE 2-PHOSPHATASE DM01656|JC1470|184-441: K123-V379 DM01656|P07953|184-442: D120-V379 DM01656|P25114|183-441: D120-V379 DM01656|P26285|182-441: D120-V379	BLAST_DOMO
			Potential Phosphorylation Sites: S22, S56, S212, S233, S302, S343, T5, T70, T185, T273, T381, Y295, Y349	MOTIFS
			Potential Glycosylation Sites: N69	MOTIFS
			Phosphoglycerate mutase family phosphohistidine signature: L191-N200	MOTIFS
7	7524757CD1	426	signal_cleavage: M1-T33	SPSCAN
			6-phosphofructo-2-kinase: R7-P206	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Phosphoglycerate mutase family: R207-I392	HMMER_PFAM
			Phosphoglycerate mutase family phosphohistidine signature: I191-Y240	PROFILESCAN
			Phosphoglycerate mutase family IPB001345A: I209-A241, V256-A268, G272-E304, E321-E366	BLIMPS_BLOCKS
			6-phosphofructo-2-kinase family signature PR00991: V82-S96, K108-I122, P134-F148, V187-S208, I209-L231, A286-P302	BLIMPS_PRINTS
			KINASE FRUCTOSE-2,6-BISPHOSPHATASE INCLUDES: ISOZYME 6PF-2-K/FRU- 6-PHOSPHOFRUCTO-2-KINASE/FRUCTOSE-2,6-BIPHOSPHATASE TRANSFERASE 2,6-P2ASE MULTI-FUNCTIONAL PD002665:A10-A64 D52-I209	BLAST_PRODOM
			MUTASE PROTEOME COMPLETE PHOSPHOGLYCERATE PGAM ISOMERASE GLYCOLYSIS BPG-DEPENDENT FRUCTOSE-2,6-BISPHOSPHATASE PHOSPHOGLYCEROMUTASE PD000730:Y210-Y323 P318-L380	BLAST_PRODOM
			6BISPHOSPHATASE TRANSFERASE 6PF2K/FRU2 6-P2ASE INCLUDES: KINASE FRUCTOSE2 MULTI-FUNCTIONAL ENZYME ISOZYME PD009472: T381-Y426	BLAST_PRODOM
			FRUCTOSE-2 SIMILAR PD114271: S252-V396	BLAST_PRODOM
			6-PHOSPHOFRUCTO-2-KINASE / FRUCTOSE-2,6-BISPHOSPHATE 2-PHOSPHATASE DM01656|JC1470|184-441: K143-V399 DM01656|JC2037|185-444: D140-V399 DM01656|P07953|184-442: D140-V399 DM01656|P25114|183-441: D140-V399	BLAST_DOMO
			Potential Phosphorylation Sites: S36, S76, S232, S253, S322, S363, T5, T90, T205, T293, T401, Y315, Y369	MOTIFS
			Potential Glycosylation Sites: N89	MOTIFS
			ATP/GTP-binding site motif A (P-loop): G26-T33	MOTIFS
			Phosphoglycerate mutase family phosphohistidine signature: L211-N220	MOTIFS
8	7516229CD1	355	signal_cleavage: M1-S48	SPSCAN
			Phosphatidylinositol-4-phosphate 5-Kinase: M1-L354	HMMER_PFAM
			Phosphatidylinositol phosphate kinases: M62-T355	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			KINASE PHOSPHATIDYLINOSITOL-4-PHOSPHATE 5-KINASE-TYPE TRANSFERASE DIPHOSPHOINOSITIDE 1-PHOSPHATIDYLINOSITOL-4-PHOSPHATE PTDINS4P-5-KINASE ALPHA PD002308:M1-F112 F112-I353	BLAST_PRODOM
			PHOSPHATIDYLINOSITOL; KINASE; DM07197|P48426|8-404:G8-Q113 Q113-T355	BLAST_DOMO
			PHOSPHATIDYLINOSITOL; KINASE; DM07197|P38994|351-756:L41-F112 Q110-I350	BLAST_DOMO
			Potential Phosphorylation Sites: S48, S150, S171, S296, S343, T18, T181, T261, T311, T325	MOTIFS
			Potential Glycosylation Sites: N46	MOTIFS
9	7516525CD1	543	Protein kinase domain: Y128-V447	HMMER_PFAM
			Serine/Threonine protein kinases, catalytic domain: Y128-V447	HMMER_SMART
			Receptor tyrosine kinase class V IPB001426: L294-K315, P316-D342	BLIMPS_BLOCKS
			Protein kinases signatures and profile: Q289-D342	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: Y303-L321, I416-V438, G350-I360, L372-D394	BLIMPS_PRINTS
			KINASE TRANSFERASE ATP-BINDING SERINE/THREONINE-PROTEIN TYROSINE-PROTEIN RECEPTOR 2.7.1.- PHOSPHORYLATION PRECURSOR PD000001:Q127-A353 G340-E453 P414-W446	BLAST_PRODOM
			KINASE ATP-BINDING TRANSFERASE SERINE/THREONINE-PROTEIN CA2/CALMODULIN-DEPENDENT BETA CG17698 CA/CALMODULIN-DEPENDENT ALPHA SERINE/THREONINE PD019141: V447-F501	BLAST_PRODOM
			KINASE ATP-BINDING SERINE/THREONINE-PROTEIN CA2/CALMODULIN-DEPENDENT TRANSFERASE ALPHA SERINE/THREONINE GLYCOGEN CALCIUM/CALMODULIN PD027014: E502-S543	BLAST_PRODOM
			KINASE ATP-BINDING SERINE/THREONINE-PROTEIN TRANSFERASE CA2/CALMODULIN-DEPENDENT BETA CA/CALMODULIN-DEPENDENT ALPHA SERINETHREONINE PD031900: M1-Q127	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004|A57156|130-399:L130-L228 Q238-V438	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN KINASE DOMAIN DM00004|P50526|136-399:E133-Q231 P247-I436	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004|P06782|57-296:I134-K166 R195-Q231 D282-V438	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004|JC1446|20-261:K129-L164 E194-Q231 V267-V438	BLAST_DOMO
			Potential Phosphorylation Sites: S69, S74, S82, S100, S117, S160, S266, S368, S457, S463, S475, S496, T26, T58, T108, T468	MOTIFS
			Potential Glycosylation Sites: N147	MOTIFS
			ATP/GTP-binding site motif A (P-loop): G523-S530	MOTIFS
			Protein kinases ATP-binding region signature: I134-K157	MOTIFS
			Serine/Threonine protein kinases active-site signature: I309-L321	MOTIFS
10	7516533CD1	445	Protein kinase domain: I30-F272	HMMER_PFAM
			Protein kinase C terminal domain: R273-I359	HMMER_PFAM
			Extension to Ser/Thr-type protein kinases: R273-A335	HMMER_SMART
			Serine/Threonine protein kinases, catalytic domain: E41-F272	HMMER_SMART
			Receptor tyrosine kinase class II IPB002011: I66-F110, I134-L185, N217-G261	BLIMPS_BLOCKS
			Tyrosine kinase catalytic domain signature PR00109: H128-L146, V194-E216, L92-E105, L236-A258	BLIMPS_PRINTS
			KINASE S6 RIBOSOMAL SERINE/THREONINE-PROTEIN TRANSFERASE P70 BETA 2.7.1.-ATP-BINDING PHOSPHORYLATION PD032092: S337-L445	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004|A53300|64-305: K55-G257 DM00004|A57459|61-302: V27-G257 DM00004|P23443|69-313: A48-G257	BLAST_DOMO
			PROTEIN KINASE C ALPHA DM04692|A37237|1-676: I50-V334	BLAST_DOMO
			Potential Phosphorylation Sites: S40, S96, S163, S295, S300, S314, S337, S341, S354, S361, S372, S399, S435, T60, T221, T310, T319, T390, Y11	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Serine/Threonine protein kinases active-site signature: I134-L146	MOTIFS
11	7516613CD1	1219	CNH domain: Y901-R1199	HMMER_PFAM
			Protein kinase domain: F25-I289	HMMER_PFAM
			Domain found in NIK1-like kinases, mouse citron and yeast ROM1, ROM2: Y901-R1199	HMMER_SMART
			Serine/Threonine protein kinases, catalytic domain: F25-I289	HMMER_SMART
			Tyrosine kinase, catalytic domain: F25-I289	HMMER_SMART
			Receptor tyrosine kinase class III IPB001824: T59-I113, W129-K168, G190-P232	BLIMPS_BLOCKS
			Protein kinases signatures and profile: W129-V182	PROFILESCAN
			KINASE SERINE/THREONINE-PROTEIN BINDING PHORBOL-ESTER ATP-BINDING TRANSFERASE GDP-GTP EXCHANGE RHO1 CDC42-BINDING PD014445:L919-S1043 F1074-S1197	BLAST_PRODOM
			KINASE SERINE/THREONINE-PROTEIN ATP-BINDING TRANSFERASE MIG-15 TYROSINE-PROTEIN 2.7.1.- PD147188:I289-P500 S795-W915	BLAST_PRODOM
			COIL COILED MYOSIN CHAIN ATP-BINDING HEAVY FILAMENT MUSCLE REPEAT INTERMEDIATE PD000002: K316-K517, Q292-Q471, Q301-Q490, L352-R569, I289-E466, Q292-R459, R358-E537	BLAST_PRODOM
			ATP-BINDING TRANSFERASE NIK KINASE SERINE/THREONINE-PROTEIN PD147187: H501-K831, E514-W915	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004|A53714|17-262: L27-S279 DM00004|P08458|20-262: V31-S279 DM00004|P10676|18-272: L27-P278 DM00004|P38692|24-266: E29-S279	BLAST_DOMO
			Potential Phosphorylation Sites: S9, S17, S77, S112, S255, S259, S264, S324, S326, S550, S554, S573, S625, S626, S633, S682, S683, S707, S721, S727, S756, S764, S880, S963, S1023, S1043, S1083, S1096, S1197, T59, T124, T187, T222, T309, T319, T351, T543,	MOTIFS
			T689, T690, T810, T816, T876, T996, T1057, Y321, Y323, Y467	
			Potential Glycosylation Sites: N33, N570, N719, N818, N1151	MOTIFS
			Leucine zipper pattern: L472-L493	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Protein kinases ATP-binding region signature: V31-K54	MOTIFS
			Serine/Threonine protein kinases active-site signature: V149-L161	MOTIFS
12	7517068CD1	1168	CNH domain: Y850-R1148	HMMER_PFAM
			Protein kinase domain: F25-I289	HMMER_PFAM
			Domain found in NIK1-like kinases, mouse citron and yeast ROM1, ROM2: Y850-R1148	HMMER_SMART
			Serine/Threonine protein kinases, catalytic domain: F25-I289	HMMER_SMART
			Tyrosine kinase, catalytic domain: F25-I289	HMMER_SMART
			Eukaryotic protein kinase IPB000719: H145-L160, Y210-G220	BLIMPS_BLOCKS
			Receptor tyrosine kinase class III IPB001824: T59-V113, W129-K168, G190-P232	BLIMPS_BLOCKS
			Protein kinases signatures and profile: W129-T181	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: M105-K118, H143-L161, S214-M236, G190-I200, W258-T280	BLIMPS_PRINTS
			KINASE SERINE/THREONINE-PROTEIN BINDING PHORBOL-ESTER ATP-BINDING TRANSFERASE GDP-GTP EXCHANGE RHO1 CDC42-BINDING PD014445:L868-S992 F1023-S1146	BLAST_PRODOM
			KINASE SERINE/THREONINE-PROTEIN ATP-BINDING TRANSFERASE MIG-15 TYROSINE-PROTEIN 2.7.1.- PD147188:I289-E648 V831-W864	BLAST_PRODOM
			KINASE SERINE/THREONINE-PROTEIN ATP-BINDING TRANSFERASE NCK TRAF2 INTERACTING VARIANT SPLICE GCK PD043898: F993-P1039	BLAST_PRODOM
			ATP-BINDING TRANSFERASE NIK KINASE SERINE/THREONINE-PROTEIN PD147187: R402-G756, S545-W864	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004|A53714|17-262: L27-P278 DM00004|P10676|18-272: L27-P278 DM00004|P38692|24-266: E29-R277 DM00004|P50527|388-627: V31-T280	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S9, S77, S112, S255, S259, S264, S275, S324, S326, S426, S446, S504, S523, S571, S580, S639, S640, S646, S647, S696, S723, S767, S776, S793, S829, S912, S992, S1146, T59, T124, T187, T222, T309, T319, T349, T467, T627, T635, T716, T750, T795, T945, T1006, Y321, Y323	MOTIFS
			Potential Glycosylation Sites: N33, N273, N333, N443, N507	MOTIFS
			Protein kinases ATP-binding region signature: V31-K54	MOTIFS
			Serine/Threonine protein kinases active-site signature: V149-L161	MOTIFS
13	7517148CD1	650	Regulator of G protein signaling domain: T54-C175	HMMER_PFAM
			Protein kinase domain: F191-F453	HMMER_PFAM
			Regulator of G protein signalling domain: T54-C175	HMMER_SMART
			Extension to Ser/Thr-type protein kinases: K454-T533	HMMER_SMART
			Serine/Threonine protein kinases, catalytic domain: F191-F453	HMMER_SMART
			Receptor tyrosine kinase class II IPB002011: L245-F289, V313-K364, D398-G442	BLIMPS_BLOCKS
			Tyrosine kinase catalytic domain signature PR00109: L271-S284, H307-L325, F417-C439	BLIMPS_PRINTS
			GPCR kinase signature PR00717: F171-N183, K230-T248, P468-I485, T493-Y506, K507-T524	BLIMPS_PRINTS
			Regulator of G protein signalling domain proteins PF00615: M15-K21, F162-K178, I270-L283	BLIMPS_PFAM
			PH (pleckstrin homology) domain proteins (P < 0.025) PF00169: S41-L47	BLIMPS_PFAM
			KINASE RECEPTOR ATP-BINDING SERINE/THREONINE-PROTEIN TRANSFERASE COUPLED BETA-ADRENERGIC MULTI-GENE FAMILY G-PROTEIN PD007430: M1-I53	BLAST_PRODOM
			BETA-ADRENERGIC RECEPTOR KINASE COUPLED TRANSFERASE SERINE/THREONINE PROTEIN ATP-BINDING MULTI-GENE FAMILY BETA ARK1PD007640: T533-Q575	BLAST_PRODOM
			BETA-ADRENERGIC RECEPTOR KINASE BETA ARK2 G-PROTEIN COUPLED TRANSFERASE SERINE/THREONINE PROTEIN ATP-BINDING MULTI-GENE PD151831: T612-L650	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN KINASE DOMAIN DM00004|P21146|193-437: V193-G438 DM00004|P32865|193-438: V193-G438 DM00004|Q09537|205-450: V193-C439 DM00004|Q09639|193-439: V193-G438	BLAST_DOMO
			Potential Phosphorylation Sites: S29, S38, S137, S156, S168, S247, S290, S343, S370, S423, S434, S487, S514, S596, S598, T187, T213, T366, T524, T533, T612, Y92	MOTIFS
			Potential Glycosylation Sites: N610	MOTIFS
			Protein kinases ATP-binding region signature: I197-K220	MOTIFS
			Serine/Threonine protein kinases active-site signature: V313-L325	MOTIFS
14	7517238CD1	603	Kinase associated domain 1: S554-V603	HMMER_PFAM
			Protein kinase domain: Y11-I215	HMMER_PFAM
			Serine/Threonine protein kinases, catalytic domain: Y11-I215	HMMER_SMART
			KINASE SERINE/THREONINE-PROTEIN ATP-BINDING TRANSFERASE ZIPPER MATERNAL EMBRYONIC LEUCINE PK38 W03G1.6 PD017644: I215-V603	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004|S52244|15-255:L13-E87 E88-M206	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004|P06782|57-296:E15-D93 E88-M206	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004|P54645|17-258:L13-E87 E88-M206	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004|S51025|18-258:L13-E87 E88-M206	BLAST_DOMO
			Potential Phosphorylation Sites: S140, S205, S308, S315, S496, S501, S600, T56, T252, T313, T339, T380, T439, T441, T470, T517, T547, T552, Y10, Y379, Y590	MOTIFS
			Potential Glycosylation Sites: N306, N437, N514	MOTIFS
			Leucine zipper pattern: L117-L138	MOTIFS
			Protein kinases ATP-binding region signature: I17-K40	MOTIFS

PF-1506 P

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
15/7517831CB1/ 1916		90040615J1	1	937
		7517831CT1	1	1916
		9689808U1	479	1163
		9689807U1	762	1440
		9689807U2	765	1648
		90040615H1	1025	1916
16/7520272CB1/ 926		95114542J1	1	584
		95114642J1	1	755
		95114758J1	1	764
		GBI_NT_008476_001.8_edit	2	925
		95114758H1	149	926
		95114642H1	218	926
17/7521279CB1/ 1382		95121315H1	1	828
		95121315J1	530	1382
		95121539H1	591	1382
		95121447J1	635	1382
		95121571H1	989	1382
18/7523965CB1/ 1678	666-1678	95141151J1	1	665
		95141159J1	1	665
		95141143H1	1	876
		GBI_NT_019583_001.8_edit	2	1677
		9773020U1	158	1087
		9773020U2	167	1047
		9746188U1	293	1057
		9773019U1	736	1536
		9746187U1	744	1267
		9746187U2	744	1526
		9773019U2	744	1593
		95141143J1	978	1678
19/7524016CB1/ 895	1-26	95183446H1	1	855
		GBI_NT_011618_001.8_edit	2	812
		95183446J1	539	895
20/7524680CB1/ 1294	1-322	95198215H1	1	819
		GBI_NT_011618_001.8_edit	2	764
		95198215J1	510	1294
21/7524757CB1/ 1354	1-22, 1333-1354	95198179H1	1	647
		95198379H1	1	710
		GBI_NT_011618_001.8_edit	2	677
		95198379J1	573	1354

PF-1506 P

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
22/7516229CB1/ 1204	1-152, 1180-1204	90046261H1	1	726
		7516229CT1	1	1179
		GBI.g20470991.edit2	300	1179
		90046261J1	349	1204
		90046277R6	396	1180
		90046269J1	474	1201
		90046277J1	475	1201
		90046293J1	497	1201
		90046169J1	568	1201
		90046285J1	582	1201
23/7516525CB1/ 1859	741-799, 1768-1859	90009054F6	1	425
		90009062H1	1	672
		90009054H1	1	701
		90009086J1	1	717
		90008970J1	1	935
		7516525CT1	1	1859
		9470229U1	205	1118
		9470230U1	728	1623
		90008978J1	859	1858
		90008986J1	888	1859
		90009070H1	974	1859
		90009078J1	1050	1858
		90008970H1	1053	1859
		90009078R6	1092	1851
		90008954H1	1114	1859
		90009094H1	1130	1859
		90009094F6	1149	1859
		90008954F6	1167	1859
		90009054R6	1267	1858
		90009277H1	1289	1859
24/7516533CB1/ 1695		90041659H1	1	767
		7516533CT1	1	1692
		9509550U1	28	1025
		9509549U1	694	1645
		90041659J1	844	1695
25/7516613CB1/ 3891	1-2632	GBI.g18553071.edit2	1	176
		90043841H1	1	805
		90043917H1	1	914
		7516613CT1	1	3891
		GBI.g18553071.edit1	231	3891
		9574496U1	613	1352
		9597296U3	642	1329

PF-1506 P

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		9597696U3	643	1438
		9597296U1	644	1447
		9597696U1	668	1289
		9628655U1	847	1618
		9628655U3	860	1623
		9628654U1	1380	2098
		9628654U3	1388	2109
		9594238U3	1541	1948
		9610640U3	1541	2000
		9610640U1	1558	1901
		9610641U3	1670	2559
		9594239U3	1872	2565
		9610641U1	1881	2589
		9597301U1	2304	3236
		9574501U1	2433	3222
		9597701U3	2438	3215
		9574501U3	2442	3223
		9597701U1	2470	3204
		9597301U3	2549	3214
		90043917J1	3007	3891
		90136633J1	3103	3891
		90043825J1	3270	3891
		90043973H1	3283	3891
		90043889J1	3337	3891
26/7517068CB1/ 3954	1-217, 3777-3954, 866-1982	90102830H1	1	603
		90102714H1	1	778
		7517068CT1	1	3954
		9580586U1	410	1315
		9580586U3	416	1127
		GBI.g18555866.edit3	440	798
		GBI.g18555866.edit2	1037	2445
		9604247U3	1132	2069
		9604247U1	1138	1925
		9604248U3	1623	2614
		9604248U1	1690	2567
		9580587U1	2379	3252
		GBI.g18555866.edit1	2446	3954
		9580587U3	2613	3234
		90102714J1	3186	3954
27/7517148CB1/ 3357	3331-3357, 1-25	90043604J1	1	655
		GBI_NT_011520_006.8_edit	26	3357
		9648674U2	377	1295
		9504918U1	447	1269
		9648674U1	580	1302

PF-1506 P

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
		9648673U1	1190	1924
		9648673U2	1194	2055
		9524840U1	1230	1956
		9524841U1	1442	2298
		9504917U1	2241	2719
		90043604H1	2545	3357
		90043712J1	2678	3356
28/7517238CB1/ 2036	2007-2036	90094269H1	1	772
		GBI_NT_008421_001.8_edit	1	2006
		9684978U2	418	1330
		9684978U1	573	1330
		9684977U2	732	1615
		9684977U1	746	1617
		90094269J1	1195	2029
		90094121J1	1199	2036
		90094213J1	1207	2029
		90094245J1	1268	2004
		90094105J1	1298	2036

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less

Table 5

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 5

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 6

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
15	7517831	142314T6	SNP00003755	149	1721	A	A	G	noncoding	n/a	n/a	n/a	n/a
15	7517831	142314T6	SNP00098537	3	1867	C	C	T	noncoding	n/a	n/a	n/a	n/a
15	7517831	142314T6	SNP00149767	132	1738	A	A	G	noncoding	n/a	n/a	n/a	n/a
15	7517831	1531602H1	SNP00023921	178	1271	T	T	G	noncoding	n/d	n/a	n/a	n/a
15	7517831	2655558T6	SNP00003755	126	1744	G	A	G	noncoding	n/a	n/a	n/a	n/a
15	7517831	2655558T6	SNP00149767	109	1761	A	A	G	noncoding	n/a	n/a	n/a	n/a
15	7517831	2829606H1	SNP00027387	110	109	G	G	A	D18	n/a	n/a	n/a	n/a
15	7517831	2836842T6	SNP00003755	140	1730	A	A	G	noncoding	n/a	n/a	n/a	n/a
15	7517831	2836842T6	SNP00149767	123	1747	A	A	G	noncoding	n/a	n/a	n/a	n/a
15	7517831	2876073T6	SNP00003755	115	1755	A	A	G	noncoding	n/a	n/a	n/a	n/a
15	7517831	2876073T6	SNP00149767	98	1772	A	A	G	noncoding	n/a	n/a	n/a	n/a
15	7517831	7758626H1	SNP00126822	384	458	C	C	A	noncoding	n/a	n/a	n/a	n/a
16	7520272	1265056T6	SNP00065601	250	838	A	A	G	G274	n/a	n/a	n/a	n/a
16	7520272	1501560T6	SNP00069832	52	916	C	C	T	noncoding	n/a	n/a	n/a	n/a
16	7520272	1501560T6	SNP00075533	128	840	C	C	T	S275	n/d	n/a	n/a	n/a
16	7520272	1968576T6	SNP00075533	204	866	C	C	T	Q284	n/d	n/a	n/a	n/a
18	7523965	1238421H1	SNP00075756	56	794	G	G	A	noncoding	n/a	n/a	n/a	n/a
18	7523965	1324236T6	SNP00033242	113	1642	G	G	C	noncoding	n/d	n/a	n/a	n/a
18	7523965	1394758F6	SNP00100133	209	491	A	A	G	M164	n/d	n/d	n/d	n/d
18	7523965	1394758T6	SNP00033242	125	1605	G	G	C	noncoding	n/d	n/a	n/a	n/a
18	7523965	1631511T6	SNP00033242	186	1543	G	G	C	noncoding	n/d	n/a	n/a	n/a
18	7523965	1964258H1	SNP00033242	56	1542	G	G	C	noncoding	n/d	n/a	n/a	n/a
18	7523965	1964258H1	SNP00136906	186	1672	C	C	T	noncoding	n/a	n/a	n/a	n/a
18	7523965	3149675H1	SNP00057801	173	785	G	G	A	noncoding	n/d	n/a	n/a	n/a
18	7523965	3149675H1	SNP00096467	177	789	G	G	C	noncoding	n/a	n/a	n/a	n/a
18	7523965	6595879J1	SNP00075756	251	798	G	G	A	noncoding	n/a	n/a	n/a	n/a
18	7523965	759508T6	SNP00033242	199	1553	G	G	C	noncoding	n/d	n/a	n/a	n/a
18	7523965	7636827H1	SNP00033242	227	1534	G	G	C	noncoding	n/d	n/a	n/a	n/a

Table 6

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
18	7523965	7636827H1	SNP00136906	97	1664	C	C	T	noncoding	n/a	n/a	n/a	n/a
18	7523965	7637976J1	SNP00075756	569	543	G	G	A	stop181	n/a	n/a	n/a	n/a
22	7516229	1329019T6	SNP00069933	170	1162	G	T	G	noncoding	n/a	n/a	n/a	n/a
22	7516229	6555450H1	SNP00023019	310	656	G	G	A	S200	n/a	n/a	n/a	n/a
23	7516525	2190612H1	SNP00128124	49	1276	A	G	A	E413	n/a	n/a	n/a	n/a
23	7516525	3780651H1	SNP00074470	124	1682	C	C	T	noncoding	0.95	0.96	0.77	0.65
23	7516525	3825922H1	SNP00074469	151	1667	C	C	T	S543	n/d	n/d	n/d	n/d
24	7516533	000364H1	SNP00002194	28	1573	G	A	G	noncoding	0.78	n/a	n/a	n/a
24	7516533	2360696T6	SNP00002194	447	1574	G	A	G	noncoding	0.78	n/a	n/a	n/a
24	7516533	2641486F6	SNP00151695	95	1482	A	A	G	noncoding	n/a	n/a	n/a	n/a
24	7516533	3078274H1	SNP00126890	34	1112	A	A	G	P350	n/a	n/a	n/a	n/a
24	7516533	3505057H1	SNP00126889	33	1081	A	A	G	E340	n/a	n/a	n/a	n/a
24	7516533	3505057H1	SNP00127060	80	1128	C	C	A	R356	n/a	n/a	n/a	n/a
24	7516533	3505057H1	SNP00127061	247	1295	A	A	G	E411	n/a	n/a	n/a	n/a
24	7516533	3505057H1	SNP00151694	133	1181	T	T	G	P373	n/a	n/a	n/a	n/a
24	7516533	4376126H1	SNP00127062	27	1369	A	A	G	K436	n/a	n/a	n/a	n/a
25	7516613	1741505T6	SNP00054334	116	2763	G	G	A	R904	n/d	n/d	n/d	n/d
25	7516613	1741505T6	SNP00124224	52	2827	T	T	C	S925	n/d	n/d	n/d	n/d
25	7516613	1852144T6	SNP00029583	89	3808	C	C	T	noncoding	n/d	n/a	n/a	n/a
25	7516613	2086173H1	SNP00029583	150	3791	C	C	T	noncoding	n/d	n/a	n/a	n/a
25	7516613	2103173R6	SNP00074035	137	331	A	A	G	R93	n/d	n/d	n/d	n/d
25	7516613	2172576F6	SNP00074035	62	327	A	A	G	K92	n/d	n/d	n/d	n/d
25	7516613	2230058H1	SNP00029582	57	2999	C	C	T	H983	n/d	n/a	n/a	n/a
25	7516613	2502887T6	SNP00029583	66	3831	C	C	T	noncoding	n/d	n/a	n/a	n/a
25	7516613	2606210F6	SNP00029582	412	2998	C	C	T	L982	n/d	n/a	n/a	n/a
25	7516613	2606210F6	SNP00124225	348	2934	A	A	G	K961	n/a	n/a	n/a	n/a
25	7516613	2606210H1	SNP00054332	20	2607	G	G	A	G852	n/a	n/a	n/a	n/a
25	7516613	2606210H1	SNP00054333	135	2722	G	G	A	R890	n/d	n/a	n/a	n/a

Table 6

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
25	7516613	2827761H1	SNP00124223	98	644	G	G	A	E198	n/d	n/d	n/d	n/d
25	7516613	3136587H1	SNP00124225	104	2935	A	A	G	T961	n/a	n/a	n/a	n/a
25	7516613	5971646H1	SNP00074036	28	1396	A	G	A	E448	n/d	n/d	n/d	n/d
25	7516613	5971646H1	SNP00074037	419	1786	C	G	C	S578	n/a	n/a	n/a	n/a
25	7516613	6203324H1	SNP00074038	314	1788	T	T	C	L579	n/a	n/a	n/a	n/a
25	7516613	7367225H1	SNP00098419	491	1864	A	C	A	S604	0.86	n/d	n/d	n/d
26	7517068	201783T6	SNP00127935	369	3700	T	T	C	noncoding	n/a	n/a	n/a	n/a
26	7517068	2836623F6	SNP00067424	105	2443	G	G	A	G807	n/a	n/a	n/a	n/a
26	7517068	2836623H1	SNP00067424	105	2442	G	G	A	A807	n/a	n/a	n/a	n/a
26	7517068	3003208F6	SNP00115029	381	640	A	A	G	Q206	n/a	n/a	n/a	n/a
26	7517068	6118733H1	SNP00115032	336	2129	C	C	T	S702	n/d	n/a	n/a	n/a
26	7517068	6448726H1	SNP00115031	288	1428	A	A	G	I469	n/d	n/a	n/a	n/a
26	7517068	6987676H1	SNP00115029	331	641	A	A	G	P206	n/a	n/a	n/a	n/a
26	7517068	7205376H1	SNP00115030	297	767	T	T	C	N248	0.65	0.49	0.87	0.64
26	7517068	7649056H2	SNP00067424	283	2441	G	G	A	A806	n/a	n/a	n/a	n/a
27	7517148	1301060F6	SNP00028255	178	2035	G	G	C	noncoding	0.99	n/d	n/d	n/d
27	7517148	1436470H1	SNP00028255	92	2091	C	G	C	noncoding	0.99	n/d	n/d	n/d
27	7517148	2008763H1	SNP00122615	72	2814	A	A	C	noncoding	n/a	n/a	n/a	n/a
27	7517148	2487070T6	SNP00122615	351	2839	A	A	C	noncoding	n/a	n/a	n/a	n/a
27	7517148	2504377T6	SNP00122615	334	2857	A	A	C	noncoding	n/a	n/a	n/a	n/a
27	7517148	2747152T6	SNP00067260	70	2411	A	A	C	noncoding	n/a	n/a	n/a	n/a
27	7517148	2836570F6	SNP00067260	348	2366	A	A	C	noncoding	n/a	n/a	n/a	n/a
28	7517238	055029H1	SNP00035691	30	998	T	T	C	S288	n/a	n/a	n/a	n/a

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☑ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☑ SKEWED/SLANTED IMAGES
- ☐ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☑ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning documents *will not* correct images problems checked, please do not report the problems to the IFW Image Problem Mailbox